Artificial intelligence in digital visual effects

Ong, Victor

2021

Ong, V. (2021). Artificial intelligence in digital visual effects. Master's thesis, Nanyang Technological University, Singapore. <https://hdl.handle.net/10356/151632>

<https://hdl.handle.net/10356/151632>

<https://doi.org/10.32657/10356/151632>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 22:55:37 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

ARTIFICIAL INTELLIGENCE IN DIGITAL VISUAL
EFFECTS

ONG VICTOR

SCHOOL OF ART, DESIGN & MEDIA

2021

ARTIFICIAL INTELLIGENCE IN DIGITAL VISUAL EFFECTS

ONG VICTOR

SCHOOL OF ART, DESIGN & MEDIA

A thesis submitted to the Nanyang Technological University in partial
fulfilment of the requirement for the degree of
Master of Arts (Research)

Statement of Originality

I certify that all work submitted for this thesis is my original work.

I declare that no other person's work has been used without due acknowledgement. Except where it is clearly stated that I have used some of this material elsewhere, this work has not been presented by me for assessment in any other institution or University. I certify that the data collected for this project are authentic and the investigations were conducted in accordance with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

11.06.2021

.....

Date

.....

Ong Victor, G1800336B

Supervisor Declaration Statement

I have reviewed the content of this thesis and to the best of my knowledge, it does not contain plagiarised materials. The presentation style is also consistent with what is expected of the degree awarded. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accordance with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

11.06.2021

.....

Date

.....

Associate Professor Benjamin Seide

Authorship Attribution Statement

This thesis **does not** contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

11.06.2021

.....

Date

.....

Ong Victor, G1800336B

Acknowledgements

I would like to thank my supervisor Associate Professor Benjamin Seide for being my mentor in the field of visual effects since my undergraduate studies in the school of Arts, Design and Media at Nanyang Technological University. In this continuous pursuit of knowledge and research through numerous sources and fascination with recent technological trends and breakthroughs with Artificial Intelligence, he has helped keep me on track instead of allowing the vast ocean of information to overthrow my direction of the dissertation. It is his personal experience in the film visual effects industry that has helped clarify some of the insights and questions which persisted through my research where I lacked thereof.

Lastly, not forgetting the overwhelming support of my family especially my mother, Maryanne Lim, who encouraged me to pursue my graduate studies, Ngee Ann Polytechnic and my students who opened my appreciation and concerns of media production and visual effects. It is the continued interest in sharing my knowledge with others in this field that led me to pursue this nature of qualification.

Contents

ABSTRACT	8
INTRODUCTION	9
CHAPTER 1 - AUTONOMOUS ROTOSCOPING.....	23
CHAPTER 2 - DEEPFAKES.....	37
CHAPTER 3 – AUTOMATED SCENE MANIPULATION	54
CHAPTER 4 - ARTIFICIAL INTELLIGENCE IN DIGITAL VISUAL EFFECTS.....	66
CHAPTER 5 – OTHER TECHNOLOGIES FOR VISUAL EFFECTS	99
CHAPTER 6 – CONCLUSION.....	111
APPENDIX A: INTERVIEW TRANSCRIPTS	122
INTERVIEW 1:	122
INTERVIEW 2:	124
INTERVIEW 3:	126
INTERVIEW 4:	127
INTERVIEW 5:	129
INTERVIEW 6:	131
BIBLIOGRAPHY:	134

Abstract

Visual effects (VFX) are today's special effects in films brought about by digitisation. Due to the increased accessible number of tools and technologies available the creation, manipulation and illusion of such effects continuously develop in photorealism and believability. Visual effects have been used to create imaginative and realistic characters, or even replace actors and backgrounds using techniques that result in an almost seamless integration that appear as if shot with the live-action elements.

Looking to the future, Artificial Intelligence (AI) development is growing. These deep learning systems can greatly assist if not, replace existing visual effects industry processes. The automation of rotoscoping in the case of Rotobot allows the reduction of industry-wide straightforward but laborious processes. Deep Video Portraits, also commonly known as deepfakes, will allow AI to process the facial movements data from a source actor to move facial features of a still portrait image of a target actor with realistic results. This dissertation investigates the possibilities when AI applications can be used in the field of visual effects - How will laborious or complex tasks be done differently to ease production time and improve efficiency or cost-effectiveness, leading to the outlook that most of the established visual effects workflows could eventually be replaced. Does having accessible AI technology in the future mean anyone outside of the visual effects profession can create high-quality film effects? It is this exploration that asks the question: How will the visual effects industry change with AI?

Introduction

Digital visual effects have been part of the filmmaking industry since the beginning of the digital revolution that happened during the 1980s towards the late 1990s. Originally, the technique and process were known as special effects ¹ but has since evolved as visual effects into today's production in films due to digitisation. Visual effects have always been very dependent on technology, and the constant adoption of technology into the visual effects process has driven rapid advancements in film production. This is possible in the visual effects industry as artists have an increased accessible number of tools and technologies capable of producing more types of filmic effects based on challenging conditions to ever integrate such elements into the film to appear as natural and seamless to realism as possible. Digitisation enabled significant advancement in visual effects by allowing production that was independent of the soundstages and movie studios, which could then be added seamlessly during post-production. "Visual effects" is "the term used to describe any imagery created, altered, or enhanced for a film or other movie media that cannot be accomplished during live-action shooting." (Okun, J. A. et al, 2010) Visual effects "involve the integration of live-action footage and generated imagery (digital effects and/or optical effects) to create environments, inanimate objects, animals and/or creatures which look realistic but would be dangerous, expensive, impractical, time-consuming or impossible to capture on film." ("Visual effects") It is "generally more cost-

¹ Special effects "are generally described as effects that can be done while the scene is being captured and are commonly called practical effects. Special effects go hand in hand with visual effects in current methodology, such that it is often difficult to determine what was a special effect and what was a visual effect. Typical special effects are bullet hits, practical explosions, rain, fire, flying rigs, motion rigs that shake sets or props or vehicles, gimbals to mimic the motion of boats or planes, and artificial ocean waves and spray." (Okun, J. A. et al, 2010 p2)

effective or practical to use visual effects than to film a scene for real, due to issues of scale or location. Visual effects if done well, are not obvious. At their best, they work to further the story being told, becoming an integral part of what makes us willing to suspend disbelief.” (Okun, J. A. et al, 2010) With affordable and convenient compositing as well as animation software available, visual effects usually include computer-generated imagery (CGI) for any individual filmmaker.

From the 1980s to the early 1990s, digital technologies were advancing across information, communication and entertainment industries alongside the emergence of the World Wide Web. This formed a dominant influence on the advancement of digital and analogue media which subsequently promoted the digitisation of analogue visual effects production technologies and various film production processes. In 1988 “audiences were excited by the use of digital ‘morphs’ in *Willow*. *The Abyss*, with its water character, was released in 1989, and *Terminator 2*, featuring a full computer-generated leading man, opened in 1991. In 1992 the first attempt was made to replicate real, recognizable creatures in a feature film – the penguins and bats in *Batman Returns*.” (Okun, J. A. et al, 2010) *Jurassic Park* demonstrated the credibility of digital visual effects in seemingly realistic interactions of photorealistic dinosaurs with actors when it came out in 1993. *The Lord of the Rings* trilogy from 2001 brought massive, simulated battles with massive digital composites using computer generation and miniatures, including a full motion captured and rendered computer-generated character. “Visual effects have allowed filmmakers to take audiences on journeys to places that have ceased to exist or that have never existed.” (Okun, J. A. et al, 2010 p10-14)

By the end of the 1990s, companies populated their rosters with young, ambitious techies who seemed unconcerned about the kinds of safeguards and benefits that Hollywood unions supplied. The majority were young white males in their twenties, many of whom were enthralled by the promise of working on projects that were redefining the visual style of film and television at the bleeding edge of screen media technology. (Curtin, M., & Vanderhoef, J., 2015 p224)

Software or computer program technologies were able to assist the artist in making crucial decisions during post-production, as well as pre-production in the form of pre-visualization or computerized storyboarding when shots require complicated camera movements or effects. The digitisation tools were readily and easily accepted into the film industry as they were user-friendly and cheaper than traditional hardware costs. The positive and adaptive integration allowed filmmaking to be cost-effective and accelerated post-production considerably during the filmmaking process. (Venkatasawmy, R. 2013) As the price of these visual effects technologies “has dropped, the industry has experienced a growth of the number of visual effects companies; quality has increased, prices have come down, and films with more modest budgets can afford high-quality visual effects work.” (Cram, C. 2012 p169-170) Working in such digital technology allowed low-cost productions to achieve high quality, stylized images with levels of craftsmanship and precision usually associated with big studio feature films. (Venkatasawmy, R. 2013 p150-151) In 1999, “Universal’s big-budget reboot of *The Mummy* had slightly over 200 shots with a visual effects component. In 2011, even the modestly budgeted comedy *Bridesmaids* exceeded that total, with around 225.” (Cram, C., 2012). Comparing to the pre-digital period of film effects, the visual effects industry has grown tremendously in film production with a high

labour cost to achieve convincing “realistic” results. Visual effects supervisor Kelly Port from Digital Domain ² cites “*Titanic*”, which he worked in 1997, had a total of 300 visual effects shots. “*Avengers: Infinity War*” (2018), he mentions in comparison, has more than 3000. (Marsh, C., 2018) Simon Stanley-Clamp, a visual effects supervisor of studio Cinesite ³, mentioned that in the 90s, there were 300 effect artists known worldwide, and today a single visual effects company would have 300 artists in the same building. (Marsh, C., 2018) This big shift in paradigm within the last three decades alone is a reminder of the technological impact digitisation has on the visual effects industry.

However, some major industry issues resulted as well. Most film audiences might believe that visual effects providers partake considerably of the tremendous box office revenues generated by blockbusters. But contracts to produce digital visual effects shots rarely include profit participation clauses that reward visual effects companies for contributing to box office success. Twenty-one visual effects companies “closed or filed for bankruptcy between 2003 and 2013.” (Giardina, C., 2014) The most publicly discussed case is on Rhythm & Hues ⁴ during 2013, the post-production house for *Life of Pi* (2012) that won the Oscar for visual effects the same year. What commonly happens is – a visual effects provider receives a script during a film’s pre-production phase which may not always include all relevant technical details at that stage – from a film producer. The process then tries to evaluate the production of each shot where visual effects are required to finally offer a fixed price to deliver visual effects for the film

² <https://digitaldomain.com>

³ <https://www.cinesite.com>

⁴ <http://rhythm.com>, recently defunct in 2020.

project as a whole. This fixed price is then pitched into a bidding competition against other local and international visual effects providers to see which one can offer to produce them at the lowest cost. If one of the visual effects providers wins the bid, they must stick to the fixed price originally pitched no matter how the film's production requirements and costs subsequently evolve. (Venkatasawmy, R. 2016 p11) As a result of relentless competition between visual effects providers trying to outbid each other by bringing down their prices to secure contracts with movie producers, visual effects artists and technicians are the ones who usually end up dealing directly with the pressures and consequences of all the cost-cutting that occurred during the negotiation and bidding stages. Under such circumstances, the bidding process fails to consider issues as the expectation in "digital visual effects compared to traditional special effects, are also seen as being nearly infinitely flexible and thus able to be revised" repeatedly. (Kaufman, 2019) It is ultimately up to visual effects producers, supervisors, designers, and artists to make it all work within the fixed budget and timeframe originally agreed upon between the visual effects provider and the commissioning movie producer. During research and development, visual effects companies are always inventing new methods and techniques. In short, one of the greatest aspects of the visual effects is the attempt to create something nobody has seen on screen before. Visual effects thus generally revolve around cost-effective techniques of removing or modifying aspects of physically filmed space, but more importantly, also attempts to introduce a spectacle. Sometimes, such effects might be obvious and extravagant, while some might be intentionally invisible. Most award-winning effects are thus appropriated to originality and technical feats in the creation of such effects. As a result, while various visual effects companies give a rough

estimate of the quotation, it is often very difficult to objectively distinguish the cost required for achieving a form of new noteworthy effect. In other words, there is always a degree of difficulty in trying to achieve something, which a studio has no idea of how to achieve in the first place. Concerns such as how difficult will it be to create an original creature or environment, or even what amount of production time will be required to make a special material react in a unique way. Thus, any form of research and development under such objectives can supposedly be considered as a sort of risk, as it is highly unlikely any visual effects company will be provided with a research and development budget. The company must carefully consider research costs in securing a reasonable bid. As a result, the least a visual effects company would do is to bid an amount that will cover this unknown cost. However, even by trying to secure this reasonable amount during the bidding to cover this specific cost, the highly competitive nature of the industry will prevent all likelihood of it happening. Since the unknown factor of such research and development is unable to concretely prove the requirements of production, a producer will more commonly often be unable to fully comprehend the factors that build-up to the actual cost of production. This lack of knowledge will often result in visual effects companies being unable to justify the costs breakdown, which includes not only the research and development but many other departments such as system admin functions, training, networking as well as archiving.

Moving forward, the visual effects industry used the latest technological features that “help artists to track projects and provide detailed analysis of the work they are doing... The developments in cloud computing are making tasks

easier and more accessible, reducing time spent on technological tasks and empowering freelance artists to access industry-leading tools from anywhere in the world." ("VFX in the modern age", 2017) The future of the visual effects industry and the available technology that can be used are amid collaborative research between companies and various research partners in an attempt to achieve better and more efficient workflows. One example would be known as "Project Sauce, a three-year EU research and Innovation project involving a nine-partner consortium, including Double Negative ⁵, Disney Research ⁶ and Foundry ⁷, that aims to greatly increase efficiency and cost-effectiveness" of visual effects. (Kaufman, 2019)

Artificial intelligence (AI) is "the simulation of human intelligence processes by machines, especially computer systems. These processes include learning - the acquisition of information and rules for using the information, reasoning - using rules to reach approximate or definite conclusions and self-correction." ("AI (Artificial Intelligence)") AI is usually adopted into a range of ongoing research and applications in special fields such as Natural Language Processing (NLP), Self-driving cars, Automation, Robotics, Machine Vision and Machine Learning. "Machine learning is the science of getting a computer to act through training and without programming. Deep learning is a subset of machine learning that, in very simple terms, can be thought of as the automation of predictive analytics." ("What is Artificial Intelligence (AI)?") The goal of machine learning is in achieving "a model (processing technique) using the training data

⁵ <https://www.dneg.com>

⁶ <https://www.disneyresearch.com>

⁷ <https://www.foundry.com>

when mathematical equations and physical laws are not promising.” (Kim, P., 2017) These models are commonly associated with neural networks. AI is a major step in today’s visual effects research and is working on changing traditional workflows to more efficient ones. At SIGGRAPH 2018, there were various discussions and presentations revolving around the various forms of AI applications for visual effects. The panel discussion delved into instances of deep learning such as convolutional neural networks ⁸(CNNs), generative adversarial networks ⁹(GANs), and autoencoders ¹⁰. There are various types of neural networks “that can be used for face simulation, fluid simulation, picture denoising, character animation, facial animation, and texture creation, among other things.” (Seymour, M., 2018)

⁸ Convolutional Neural Networks (ConvNet/CNNs) is “a deep neural network specialized for image recognition. It is an old technique, which was developed in the 1980s and 1990s. However, it has been forgotten for a while, as it was impractical for real-world applications with complicated images. Since 2012 when it was dramatically revived, it has conquered most computer vision fields and is growing at a rapid pace.” (Kim, P, 2017 p121)

⁹ Generative Adversarial Networks (GANs) are a type of generative modelling that employs deep learning techniques like convolutional neural networks. In machine learning, generative modelling is an unsupervised learning task that involves automatically discovering and learning regularities or patterns in input data so that the model can be used to generate or output new examples across a variety of problem domains, “most notably image-to-image translation tasks such as translating photos of summer to winter or day to night, and in generating photorealistic photos of objects, scenes, and people that even humans cannot tell are fake.” (Brownlee, J., 2019) It is a system with two parts - a discriminator network and a generator network. “The generator and the discriminator networks are constantly competing against each other” (Wang, L. et al, 2020). The generator’s task is in creating data a discriminator will be unable to discriminate from real data. In short, the generator tries to form an estimate to the true data on an image, while the discriminator classifies the data passed onto it as real or fake.

¹⁰ Autoencoders “are an unsupervised artificial neural network that learns how to efficiently compress and encode data then learns how to reconstruct the data back from the reduced encoded representation to a representation that is as close to the original input as possible.” (Badr, W., 2019) To specify a little more, it is made of an encoder, a decoder and a bottleneck. It is an unsupervised learning algorithm that, tries to learn the identity function and give an output as close to the input as it can.

Source: <http://www.andrew-whitehurst.net/pipeline.html>

Fig. 1. Stages in a typical visual effects pipeline.

However, AI systems are not included in any visual effects pipeline ¹¹. As a result, the implementation of the tools may be difficult for artists even if there is the availability of open-source codes. There “should be the integration of deep learning libraries as part of the core visual effects pipeline, including the automatic collection and processing of data of such tools. These tools need to be more flexible, controllable and adjustable for creative tasks, and the artists and technical people need to be more prepared to use and adjust these technologies in their workplaces”. (“Interview: Dmytro Korolov on the results of SIGGRAPH’s AI tools for VFX panel”, 2018) It will take a long while before neural networks

¹¹ A visual effects pipeline is a workflow commonly practised in the visual effects industry during the production of a film. Refer to Fig. 1 for the diagram of the pipeline. Details on the workflow will be covered in the beginning of Chapter 4.

eventually become integrated into the visual effects pipeline. AI technology that visual effects houses would invest to use effectively is unfortunately not as widely comprehended as common technical knowledge. When issues arise in the workflow, it can be difficult to determine whether the problem was caused by a flaw in the data, the model configuration, a flaw in the open-source code, a flaw in a proprietary revision of that code, or even a fundamental misunderstanding about how the core source code approaches a specific problem. (Anderson, M.). A persistent potential for wrong interpretation by the AI systems which are insufficiently or incorrectly trained might prove unreliable and troublesome. Because knowing what happens inside a neural net is still a work in progress, many have claimed that this request has conceptual and practical issues. (Wachter, S. et al., 2017)

The majority of the programming of such systems are unfortunately not directed specifically for visual effects practises in the first place. As a result, the varied ways of adoption towards visual effects solutions are uncertain in the AI networks predictability of performance. The various research system technology covered in such developments usually result from “the world’s wider-ranging research groups: massive collective efforts that are seeking to make self-driving cars which don’t crash, advertising insights that work, robots that can replace actual physical workers, and security systems that can track individuals by characteristics.” (Anderson, M.). Autonomous and intelligent labour is something visual effects companies would be interested in investing in. However, for AI systems and machine learning to be a commonly understood and used solution, there needs to be some sort of collaborative effort that involves various inter

studio cooperation to develop some type of open-source platform that is shared globally, such as “Pixar Animation Studio’s ¹² Universal Scene Description (USD) format; continuing development on the OpenVDB C++ library (computer language) initiated by Dreamworks ¹³; the OpenColorIO (OCIO) colour management solution; the interchangeable computer graphics file format Alembic, developed by Sony Pictures Imageworks ¹⁴ and Industrial Light & Magic ¹⁵ (ILM); Sony Imageworks’ Field3D voxel data storage system; and the OpenEXR projected initiated by ILM.” (Anderson, M.) If visual effects companies resort to relying solely on the current most popular implementation of AI systems or tools for production, then there is a high chance of potential incompatibility and wasted investment when newer and more feasible states of updates from open-source networks and libraries become available. For example, every time the source platform for machine learning makes a version leap in release, whatever in house visual effects tool developed from such systems would likely have issues or become unusable. The companies will then have to modify or re-program the workflow and tool to fix everything - an issue that will likely persist with any subsequent iteration of the updated technology. The rather unilateral method that mainstream research is feeding innovation in AI-driven visual effects from the world's economic and state-sponsored powerhouses would recommend stronger development and version control, a crucial feature of collaborative open-source AI

¹² <https://www.pixar.com>

¹³ <https://www.dreamworks.com>

¹⁴ <https://www.imageworks.com>

¹⁵ <https://www.ilm.com>

efforts in the sector. (Anderson, M.) As soon as these problems have been lifted, the potential change in the visual effects industry will be revolutionary.

This research will identify possible AI developments that seem to be potential candidates to assist digital visual effects artists to ease production time and improve efficiency or cost-effectiveness. More specifically, the disciplines or fields that I perceive as often being more labour intensive or otherwise complicated in production such as rotoscoping, computer-generated characters or creatures, difficult or unforeseen production schedules due to unfavourable locales or conditions, certain post-production edits and clean ups of unwanted obstructions or elements, reconstruction of lightings and set extension, are the departments in visual effects I would like to address with the possibilities AI can perhaps assist or replace. Chapter 1 will be the introduction of rotoscoping, and the investigation of various AI processes that enable automated features to potentially assist the laborious endeavour. Chapter 2 is the concept of synthesizing facial features and consequentially resulting in a new take on identity infringement via deepfake, as well as exploring the powerful application of AI in helping in character creation or even replacements under the same context. The general importance and practice of digital set extensions for world-building in various scenes in films are introduced in Chapter 3, with a look into AI that will greatly assist in environmental mood and aesthetic translations during post-production edits. Some instances of the AI systems presented also assist in laborious background clean-ups, which will prove invaluable to post-production efficiency as well. In Chapter 4, I will attempt to predict a future forecast of the state of a new visual effects industry workflow that will possibly persist with the

successful inclusion of AI induced techniques, given that artists have full control and error-free assistance from the developing technology. Such AI systems will be presented in various stages of implementation of the typical visual effects pipeline, spanning the next decade while I attempt in discussing certain methodological adoptions and changes along the way. The section will also cover my prediction on how the AI technologies might gain entry into the visual effects industry by presenting four distinct scenarios that access digital software trends, real-world influences, and potential sources of development that involve AI technology in various aspects and applications. In Chapter 5, I will introduce some miscellaneous yet video affiliated research which are possible future enhancements to the video medium format, but not directly associated with the visual effects techniques. It will also cover a cross-analysis on my research, comparing the proposed future workflow to another series of visual effects adoption happening concurrently known as the Virtual Production (VP) - where a huge emphasis on pre-production techniques and preparation is done for the actual film shoot production that aims at reducing post-production processes with in-camera footage effects. I will also conclude the research by discussing the general view of visual effects during the next step of digital technology, and how it and the varied AI assistance might also penetrate other media-related fields of commerce as a common commodity and visual language. In placing the AI techniques as an alternative to some current visual effects industry's practices, I hope to investigate the possibilities these new "tools" will impact towards the visual effects pipeline, should it remain somewhat similar in its design or otherwise, resulting in a different state and workflow of the visual effects industry. I have also conducted a minor survey to investigate how visual effects

veterans of today's industry feel about AI, and their general perspectives on the feasibility and acceptance of some AI methodologies into the industry practices. There is a total of 6 surveys listed at the end of this paper, under Appendix A, and they provide reasonable insight into the current visual effects hopes and challenges as spoken by the artists themselves. (As of the year 2019) Although there are some instances of successful implementation of minor AI assistance and assistance in the visual effects framework, the current development of most known AI systems and applications is by no means of a professional and industry state ready to be utilized in actual film production. One can only imagine the drastic change when they are commercially available to both consumer and professional-grade licenses as a fully functioning and easy to use solution.

Chapter 1. Autonomous Rotoscoping

Rotoscoping is a technique in the animation and visual effects industry used for decades since the invention by Max Fleischer in 1917. It involves a laborious process of projecting a pre-recorded moving image of actors onto a glass panel in which the artist traced manually with a pencil over the image. The “projection equipment is referred to as a rotoscope” (“Visual effects”), and this technique allowed realistic fluid motions in the animations. On projection, the number of frames would usually take a full year to film, trace, composite with backgrounds, and filmed again. The result is usually just a few minutes of animated footage. (Cartwright, L. 2012) From the 1940s through to the 1960s, rotoscoping started to be used primarily for matte painting for films. It involved “tracing the elements that had to be isolated onto an animation cel. The outline traced onto the cel then would be filled in with black paint, so that it would block the appropriate section of the frame.” (Seymour, M., 2011) In Alfred Hitchcock’s *The Birds* (1963), the film required a shot where scores of birds attack a small town. As the nature of the shots included white birds shot against cliffs, beach or the sea, it created a great degree of difficulty in isolating them with former techniques such as the Bi-pack process matte filming¹. Resorting to rotoscoping process for “500 frames in one shot of the attack took two artists 3 months to complete. In *Return of the Jedi* (1983), animated walking weapons known as the *All-Terrain Armoured Transport (AT-ATs)* were *matted* into real forest scenes.” (‘What is Rotoscoping?’) In this case, the trees were traced and isolated frame by

¹ Bipack Matte Filming – “The process of loading two reels of film into a camera, so that they both pass through the camera gate together resulting in various techniques of in-camera effects.” (Read, P., et al, 2000.)

frame. This allowed the animated machines to move between trees for added realism in the scenes.

Source: <https://www.fxguide.com/fxfeatured/the-art-of-roto-2011/>

Fig. 2. A breakdown of roto work in isolating elements of a scene as mattes within a single frame.

The visual effects industry adopted this technique to manually create mattes ² for elements “on a live-action plate, so they may be composited over another background.” (“Visual effects”) These mattes can be used to replace a real prop with a computer-generated element, or even apply selective image filters over parts of a video frame. (Agarwala, Hertzmann et al. 2004) It is one of the most essential techniques used in almost every single visual effects shot. In the 1930s, chroma photography was still in its infancy. It was referred to as “travelling matte” at the time. (“The Benefits and Drawbacks of Chroma Shooting for VFX”, 2018) The use of a blue screen as a background behind an actor during

² Mattes in film and television is used to separate an element from the background and then treat the separated element differently, i.e. with a filter or colour correction, OR to composite a foreground element over a different background.

filming allowed the use of optical filtering. These mattes are generally less intensive to create rather than rotoscoping due to the bright blue colour. When running the filmed blue through a red filter, the result is a black opaque region on the frame. As such, using high-contrast black-and-white film to create the mattes, the silhouettes of the actor can be created. These silhouettes allow the film to be rewound and re-exposed to lay the actor into the “hole” which the matte creates during overlaying.

Gradually, through technology and various research and development, this technique was upgraded when films would require set extensions or computer-generated elements incorporated into shots. It is a solution that allows a new element to be superimposed over a background. Today, the technique has changed to be more efficient and affordable known as chroma keying – commonly used during compositing when attempting to layer various images or videos into a single scene based on colour hues. In recent modern filming, visual effects departments often work alongside live-action shooting in setting up green screens. Green screens are essentially long blankets of fabric fixed or fastened tautly to a stand, composed of nylon to prevent creasing and folding. Green screens can be as little as a modest background screen set behind a single subject or as large as a complete soundstage coated from top to bottom. Green screens can also be used to hide objects such as props or objects that an actor can interact with, which are subsequently digitally recreated in post-production. Because the chroma key software operates on the idea of removing a single colour from the frame, that colour should be unusual. The vivid, highly saturated green colour utilised in green screen production is often distinct enough that it is not found in

clothing or items, and it is the primary reason for its clearer delineation from natural elements, particularly under bright lighting. (Stojković, N.,2020)

Therefore, the colour green is most suited to be used for shots that require outdoor lighting. The blue colour is now used typically for night shots to eliminate the unnatural “bleeding” green colour that is commonly reflected onto the surface of the subjects. However, chroma-keying sometimes still leads to unwanted colour spills onto actors which may require manual clean up. Often as most of the filming is done in a studio setting that is disparate from live-action on-set conditions - there is a greater degree of difficulty in doing perfect lighting that matches with the painted backgrounds. Rotoscoping is still required to clean up unwanted sections of chroma keyed elements, albeit to a much lesser degree of finesse and details. (Mozar, A., 2016)

Modern rotoscoping is mostly done digitally using programs such as After Effects, Silhouette, Nuke as well as Flame/Smoke. The transition to digital-based rotoscoping began as early as the 90s because of a particular software called Colorburst. Colorburst is a kind of digital image editing tool like what we have commonly known to be Photoshop, which eventually became Matador. As a result of this, in one-quarter of the time, eight or more rotoscope artists have been reduced to requiring only one rotoscope artist in achieving similar production results. The computer could use a single pre-drawn frame as a base template, meaning that the tedious work of continuously redrawing frames is no longer required. Various rotoscope tools function using a series of vector points that are connected resulting in a line or curve, known as splines. These can be adjusted

frame by frame with slight deformations and movement according to the traced object.

Rotosplining “is the process of creating vector shapes to manually cut an element out of its background. The position of these shapes can be adjusted on individual keyframes, while the software can interpolate a transition between the frames.” (Seymour, M., 2011) It is this automatic interpolation that significantly reduces the number of manual adjustments required. Most rotoscoping will allow an artist to create multiple rotosplines that can be adjusted and keyframed individually. By doing this, certain more complex sections can be selectively removed from images. For example, “an actor running would have separate shapes for body regions such as the hand, forearm, upper arm, chest, torso, thigh, shin, etc. By breaking the shapes down into smaller elements, it is much faster to set the keyframes by moving the shape and not individual points, and the software will interpolate much more accurately. When doing complex rotoscoping, such as a jungle with a character running through foliage or something similar with many moving independent motions, it is vital to be able to play multiple shapes in real-time over the background footage.” (Seymour, M., 2011) It has gotten more demanding in current films, where such complex rotoscopes meant longer frames, as well as a higher degree of accuracy needed in larger resolution footage for computers to process and the artists to adjust.

Rotoscoping services are still mostly required in most films and have been readily outsourced to visual effects houses located in India and China for their affordable manpower. Their services generally include the creation of rig

removal, reflection removal, colour grading, as well as rotoscoping isolation. etc. But over time in these countries, “the expertise and skill set of their local visual effects artists have gradually developed, they have acquired more experience, and they inevitably have to be paid higher salaries.” (Venkatasawmy, R. 2016) This “specialized role, as even in the developing economies, are charged out at around a hundred dollars per day. Even if an artist is capable of rotoscoping at a rate of one frame per minute, one second of footage will take 24 minutes to process, meaning that in one day, an artist can process 20 seconds of footage featuring one instance of a person. Thus, two people in the footage mean twice the amount of work, costing at the very least two hundred dollars and requiring two trained artists to work concurrently and have the cost of two interactive licenses of digital software.” (Hodge, S., 2018)

Currently, computers can “hunt for humans and automatically rotoscope, albeit not on production level yet.” (Kaufman, 2019) There are developments and implementations from within the visual effects industry itself currently. Visual effects companies like Digital Domain and Industrial Light & Magic shared that rotoscoping is a great example as the basic AI tool can already do garbage matting³ and their AI development is “close to solving a deep learning-based rotoscoping solution that will be just as good as an artist.” (Seymour, 2018) Arraiy Inc.⁴ is a company that demonstrated its rotoscoping system’s capabilities during “*The Silicon Valley Deep Learning Group*” conference in 2018 by showing “a human

³ Garbage matting is the term referred to “the use of mattes to identify regions of an image that should be ignored. Typically, loose garbage mattes are created to remove rigging or other obvious regions from a green screen image.” (‘Garbage Matte.’, 2019)

⁴ <https://www.arraiy.com>

being extracted from a background studio in real-time without colour keying. The company trained the system by inputting vast amounts of specially shot footage, recreating in the process some of the most difficult problems in background extraction” – producing high-quality mattes around strands of hair. (Anderson, M., 2018) The development of such rotoscoping networks is currently still being pursued and refined.

The main development of this technology stems from computer vision of the machine learning AI research that deals specially with images and segmentations. A common term used is “image segmentation” which refers to “dividing an image into groups of pixels based on certain criteria. A segmentation algorithm takes an image as input and outputs a collection of regions which can be represented as a collection of contours or a mask where each segment is assigned a unique grayscale value or colour to identify it. When we group pixels based on colour, texture, or other low-level primitives, we call these perceptual groups superpixels.” (Mallick, S.,2018) Every pixel on the screen of an image can be labelled with a class and we call this semantic segmentation ⁵. This is the most commonly used technique in the majority of autonomous rotoscoping algorithms. However, “we can say which pixels belong to the class ‘person’ by simply checking if the mask colour is similar at that pixel, but we cannot say if two similar coloured mask pixels belong to the same person or different ones.” (Mallick, S.,2018) Semantic segmentation comes into many fields of AI-related research such as autonomous driving for vehicles, facial segmentation that is giving rise to

⁵ Please refer to Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I. published in 2016.

mobile apps as well as virtual make-overs, indoor objects segmentation commonly used for “Virtual Reality (VR) and Augmented Reality (AR) applications as well as geo sensing applications through satellite imaging.” (Chakraborty, A., 2019)

Autonomous rotoscoping networks are currently also available for the public domain to access and use. OpenCV or known as Open-Source Computer Vision Library is “an open-source computer vision and machine learning software library that was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in commercial products. Along with well-established companies like Google, Yahoo, Microsoft, Intel, IBM, Sony, Honda, Toyota that employ the library”⁶ to do AI-related tasks, it can be programmed to perform visual effects rotoscoping tasks via computer languages such as Python. At this moment, the main limitations of this framework are that changes and modifications have to currently be done with the programming language as a third party software and there is no ease of control via User Interface (UI) elements within a paired compositor software used to achieve the varied post-production effects required. As a result, the additional steps and programming knowledge as a pre-requisite measure for the development of rotoscoped elements to be used for post-production is unreliable and overly complicated to fix to a reasonable and acceptable degree thus far. The majority of the visual effects artists who had experience using AI-assisted rotoscoping networks or algorithms have mentioned this lack of control and quality as seen in Appendix A. As a result, AI neural network-assisted rotoscoping

⁶ Taken directly from OpenCV’s official website, <https://opencv.org>

has been commonly utilized for a simpler production process that does not require the finesse and fidelity of a proper rotoscoped solution.

Sam Hodge, founder and creator of Kognat ⁷ mentions that “artists requiring the bash roto ⁸ to complete their tasks are unable to start the work until the masks are available for the footage. As a result, often work is completed quickly in a substandard way, decreasing the quality to release a task for downstream departments. The cost of a bash roto can be as much as forty per cent of the total cost of the final rotoscope. Often the bash roto cannot be reused to form the foundation for the finished masks, so in this way, it is throwaway work. The cost of bash roto for typical visual effects shot can be 2 to 60 hours of labour depending on the length and complexity of the shot and number of elements that need masks.” (Hodge, S., 2018) Through image segmentation, where the network aims to detect and consistently monitor specific features in a video stream, neural networks can now extract distinct multiple items from a background. (Anderson, M.)

A research technology paper “*Background Matting: The World is Your Green Screen*” by Sengupta, S., et al. has a proposed methodology using AI deep network for creating mattes that separate a person in the foreground from the background scene by taking two shots of a scene – one with and one without the person leaving only the background. The imaging process requires the user “to step out of the frame after the shot to capture the background and works with any

⁷ <https://kognat.com>

⁸ In the article “Rotobot, bringing Machine Learning to Roto” from fvguide, “Bash Rotos are quick and simple rotoscopes that allows artists to check or visualize work, before committing to more costly and time-consuming high-quality rotoscope. They are often used in ‘slap’ composites which allow artists and supervisors alike to explore ideas and test approaches.” (Seymour, M., 2018, November 30).

camera with a setting to lock the exposure and focus.” From the input, the system “extracts a soft segmentation of the subject and can additionally utilize nearby frames to aid in matting.” (Sengupta, S., et al, 2020) It then also applies a deep adversarial loss prediction algorithm to synthetically recover lost matte and colour.

Source: *Background Matting: The World is Your Green Screen* by Sengupta, S., et al., 2020.

Fig. 3. Example of the system process and resulting composite image.

Despite showing promising high-quality foreground and alpha mattes, the system still overly relies on the constraints of pre-deep learned data of backgrounds and minor camera motions. It requires “a small amount of foresight in shooting a scene but is far less time consuming” (Sengupta, S., et al, 2020) than standard rotoscoping requirements. For example, while shooting a scene on set, it will simply require a first still or controlled camera movement of the locale without any actors filmed separately before letting them enter the shot and beginning the actual take. That is - for every scene that requires the automated matte where tricky effects or set extensions are needed in the shot, the post-production artists will have a background set with no actors, and one with everything on set. There is, however, also a system that does not require this additional step.

Rotobot by Kognat is a plugin that can extract multiple elements from footage via this means, and its automation currently allows a substantial amount of cost saved for any visual effects company. It is still in development and consumers who purchase a license are still receiving numerous updates. Currently, it is compatible with a few visual effects industry software such as Autodesk ⁹ Flame 2020, Blackmagic Design ¹⁰ Fusion Studio, Nuke by Foundry, Blackmagic Design DaVinci Resolve as well as Natron by INRIA ¹¹. By incorporating Rotobot into their current visual effects production, some companies could have a turnaround time “condensed from a 24-hour turnaround to a 5-minute calculation on a sufficiently scaled render farm with appropriate licenses, meaning that downstream artists can start their tasks earlier and there can be more parallel workflow where the roto artists are spending time that would have previously spent on bash roto finishing the final quality work, while the effects or lighting artist are working on the elements of the shot that will integrate with the roto mask layer.” (Hodge, S.,2018) Rotobot’s speed of computation is only limited by the scale of computational resources occupied and the number of concurrent licenses purchases.

⁹ <https://www.autodesk.com>

¹⁰ <https://www.blackmagicdesign.com>

¹¹ <https://www.inria.fr/en>

Source: <https://kognat.com/higher-resolution-masks-demo/>

Fig. 4. Example Rotobot auto- segmentation output of selecting “person” from various other choices such as “aeroplane”, “bicycle”, “bird”, “boat”, “bottle”, “bus”, “car”, “cat”, “chair”, “cow”, “dining table”, “dog”, “horse”, “motorbike”, “potted plant”, “sheep”, “sofa”, “train”, “tv”.

The capabilities of running autonomous rotoscoping simultaneously with little to no breaks in the process have given Rotobot some considerations to studios around the world. The software is accelerated by NVIDIA’s CUDA ¹² computation on GPUs that are routinely used in production since its launch in 2006.

When these autonomous rotoscoping technologies are proven to be working at a final quality standard, the industry’s leading grunt work of laborious

¹² Based on Nvidia’s developer website <https://developer.nvidia.com/cuda-zone>, The first GPUs “were designed as graphics accelerators, becoming more programmable over the 90s, culminating in NVIDIA’s first GPU in 1999... Ian Buck later joined NVIDIA and led the launch of CUDA in 2006, the world’s first solution for general-computing on GPUs. Since its inception, the CUDA ecosystem has grown rapidly to include software development tools, services and partner-based solutions. The CUDA Toolkit includes libraries, debugging and optimization tools, a compiler and a runtime library to deploy applications. CUDA accelerates applications across a wide range of domains from image processing, to deep learning, numerical analytics and computational science.”

rotoscoping requirements will be of the past as artists can easily and quickly generate ready mattes without spending hours of manual labour. Visual effects companies also require fewer appropriate operating licenses to be able to generate mattes for production. This might result in a worldwide reduction in demand for rotoscoping artists, affecting dramatically the labour force from visual effects companies located in India and China. As a result, it would be most likely that artists will be able to exercise their exploration and creative freedom to a larger extent with the availability of time, and visual effects companies will be inclined to provide quick and high-quality final renders allowing the film productions to be well on schedule. By bypassing the need for rotoscoping labour, artists from various visual effects departments can start on the other effects while using the final quality rotoscope already available from the very beginning which significantly reduces the chances of errors caused by unforeseen issues during manual syncing of rotoscopes and updates from bash rotos. However, these networks and systems will hopefully have the means to perfectly segment various elements on the footage, including allowing the ability to edit and customize the auto-generated mattes to the user's preference and requirements. As an example, to be able to differentiate people from the background, as well as various objects, and certain background areas. The ability to select each corresponding segment will give the artists a full range of control, and the choice to keep or remove the selection. The role of a rotoscope artist will not completely disappear per se, as there will still be a required rotoscope editor who will oversee the sorting and controlling of the generated mattes to be used in the respective departments.

These examples show how AI can potentially change or remove inefficient and uncreative processes from the visual effects production, resulting in a more reliable toolset that can be used for a significant reduction of labour and time requirement.

Chapter 2. Deepfakes

Commonly known as deepfakes amongst public media, these new forms of media modification have been utilized in a few instances of video related editing in pornography as well as fan-based communities of cinema pop culture and public controversies. It was created in 2017 by a user on a popular online forum platform Reddit named “r/deepfakes”, who “used Google’s open-source deep-learning library to swap pornographic performers’ faces for those of actresses.” The codes that “deepfakes popularised globally today are mostly descended from this original code” that was released. (Adee, S., 2020) This deepfake technology can effectively replace anyone in any recorded image or video which they did not participate in with the assistance of an AI network. In films, such replacement requirements have existed in the visual effects industry for decades. For example, in the film *Fast & Furious 7* (2015), late actor Paul Walker still appeared on screen with the help of computer-generated imagery. However, the production process would take Weta Digital ¹ studios close to a year to create the result released in the film. In the end, there was a shot of a car jumping between Abu Dhabi's Etihad Towers which consisted of roughly 260 shots involving one of the deceased actor’s brothers who was facially replaced by computer-generated faces of Paul Walker’s. There was no need to develop a computer-generated body as the brothers closely resembled his. “An additional 90 or so shots were completed with outtakes or older footage of Paul Walker that had already been shot and which were then relit and repurposed digitally, resulting in the shots of Paul

¹ <https://www.wetafx.co.nz>

Walker driving in a car alongside actor Vin Diesel in the film's closing moments.”
(Giardina, C., 2015)

A variation of such an example would be character replacement, duplication or creation techniques that have been a major visual effects problem for animation and filming. Industrial Light & Magic modified Ariana Richard's stunt double's face on-screen with a computer-generated facial replacement of Ariana Richard's during the air vent escape scene when escaping from the raptors in *Jurassic Park* (1993). In *The Crow* (1994), the first computer-generated imagery recreated deceased actor Brandon Lee, such as Lee's face on the body double such as the shot of Draven falling from the window complete with simulated blood, and full shots on another background. *Tron: Legacy* (2010) used visual effects to create an advanced photorealistic computer-generated actor. The computer-generated version of “young Jeff Bridges is constructed from footage of him in *Against All Odds* (1984) made 2 years after the original *TRON* (1982). This was then retrofitted onto motion-capture² performance by Bridges himself and stunt double John Reardon who was his stand-in. The whole thing took approximately 2 years to complete.”³ Currently, finding the placement of a person in footage is a common task for “applications such as 3D avatar creation, facial animation transfer, video editing and compositing.” (Genova, K., 2018) The conventional approach for character replacement is to search the space through the tracking software and motion capture suits and use computer-generated techniques to

² Based on <https://whatis.techtarget.com/definition/motion-capture-mo-cap>, Motion Capture (Mo-Cap), is the use of sensors on real bodies (objects and beings) to track motion and record the movements digitally, typically for transferring to virtual 3D imaging.

³ Description sourced from <https://computeranimationhistory-cgi.jimdofree.com/tron-2010/>

generate an animated 3D face or character image that “matches the character placement by optimizing shape, texture, pose, and lighting parameters. This approach requires a complex, non-linear optimization that can be difficult to solve in practice” (Genova, K., 2018), and over the years the details and realism in creating believable photorealistic characters are recent achievements in the visual effects milestones. Such computer-generated imagery techniques often require “a large set of artistic and technical skills, ranging from sculpting, painting and animation to algorithm design, math, physics and biology. The artist takes into consideration the underlying bone and muscle structures, which are animated under the constraints of what real human faces are capable of.” The model is then “textured to give the skin a realistic colour and surface appearance, which includes complex algorithms calculating how light should reflect off the skin or scatter underneath the skin in the softer parts of tissue, such as the ears or the nose. It needs to be lit by lights in 3D space to match the lights in the real scene and after rendering the model into an image sequence (a video) a compositing artist has to combine it with the original footage of the real actress or stand-in.” (Harris, S. A. K., 2019) The capabilities of the computer-assisted techniques are an example of the technology that has developed and enabled various avenues in modifying or changing the way filming processes were formerly done as far as character effects were required, where pre-computerised techniques often relied solely on camera tricks or practical make-up effects.

There is only one area of scrutiny in visual effects where only absolute perfection is acceptable – realistic human faces. People can scrutinise faces to an incredible degree and when visual effects attempt to create human faces, we can

experience what is called the uncanny valley⁴. Visual effects artists struggle to remove the negative experience as computer-generated characters become more photorealistic. Subtle nuances in facial features or bodily appearances prove difficult to fool the human eye and the audiences' perception of truthful realism, as most people have grown accustomed to similar visions through everyday life to give reasonable deductions, allowing them to apply prior evaluations accordingly. Even *Star Wars: Rogue One* (2016) which has some of the most realistic digital humans ever created in visual effects history failed to be perfect in this aspect. To have a popular character named *Grand Moff Tarkin* appear on screen, Industrial Light & Magic tried to strike a balance between a digital figure who seemed realistic and one who looked close to late actor Peter Cushing. For example, the lighting in *Star Wars: A New Hope* (1977) was lit differently than *Rogue One* raising questions of how to adjust the lighting on the character. If the digital figure were lit similarly in *A New Hope*, it would improve his likeness as Tarkin but worsened the sense of realism since he would look different from other actors in the scene. Peter Cushing's subtle nuances of talking also proved difficult to replicate, such as when he "makes an 'aah' sound, he doesn't move his upper lip. He only opens his jaw about halfway, and makes this square shape with his lower lip, that exposes his lower teeth." (Itzkoff, D., 2016) The team eventually decided to pay more attention to realism than likeness, resulting in a figure that seemed like a relative to Peter Cushing instead. This drew numerous criticisms. The Huffington Post called it a giant breach of respect for the dead (Thompson, Z.,

⁴ "Uncanny Valley" is a term derived by Masahiro Mori, a robotics professor at the Tokyo Institute of Technology, on how he envisioned people's reactions to robots that looked and acted almost like a human. In particular, he hypothesized that a person's response to a humanlike robot would abruptly shift from empathy to revulsion as it approached, but failed to attain, a lifelike appearance. This descent into eeriness is known as the uncanny valley. (M. Mori, et al, 2012)

2017) and The Guardian described it as a digital indignity. (Shoard, C., 2016) One exception is *Blade Runner 2049* (2017) that fully recreated actress Sean Young for the character, Rachael, from the original movie *Blade Runner* (1982) using computer-generated imagery techniques. However, this almost perfect replication process took The Moving Picture Company (MPC) ⁵ visual effects team over a year to produce the believable computer-generated digital double which appeared for 2 minutes in the film. (Montgomery, J., 2018) In considering the requirements of visual effects to successfully produce an acceptable quality of character replacement in this sense, the industry does seem to fall short in demonstrating reasonable solutions. Now, deepfake technologies developing from fields of automatic computer graphics or machine learning systems can quickly generate and produce videos or images to a degree that is worth investigating.

There has been a lot of confusion around the term “deepfake,”. It is something that may be regarded as misleading and pertains to the undesirable nature of misuse of such a technology. Unfortunately, the term has become a catchphrase or singular class to “describe everything from state-of-the-art videos generated by AI to any image that seems potentially fraudulent” (Adee, S., 2020) and often seemingly illegal. The main system that drives deepfakes is essentially AI machine learning, which allows replacements of character faces at a much lower cost. To create such a deepfake, the system needs to be trained within a certain amount of time ranging from hours to days with data of the same or even different specified character in various angles and light conditions. This results in a neural network that can distinguish and produce a realistic and accurate

⁵ <https://www.moving-picture.com>

comprehension of the intended effect. After which, the trained network will attempt to digitally alter the character into the different character that is trained with. Despite the assistance of AI system networks that are faster compared to prior computer-generated techniques for facial replacement, it still requires a lot of time to produce realistic and desired results. An artist must still make varied changes and provide correct training parameters to reduce as many problems and artefacts produced, and as such, the entire process is relatively complex and has no functional state to be used directly within current visual effects proprietary software. Many people believe that generative adversarial networks (GANs), a type of deep-learning algorithm mentioned at the start of this paper, will be the major engine of deepfakes development and research. (Adee, S., 2020) The technology thus far has garnered plenty of fascination in science and the general public. MIT-IBM Watson AI Lab ⁶ created an app that allowed the transformation of photographs to ink portraits, oil paintings, faux watercolours, resulting in widespread excitement. Generative adversarial network models in the app “combines opposing neural networks to produce believable outcomes – a generator that looks at examples and tries to copy them, and a discriminator, which deciphers if they are real by comparing them with existing images... scanned over 45,000 portrait images to train the program, including paintings by Van Gogh and Rembrandt.” Samsung AI lab ⁷ also utilised similar neural networks to make the “Mona Lisa ‘smile’ and created interactive portraits of Salvador Dali and Marilyn Monroe from a single image.” (Feinstein, L., 2019)

⁶ <https://mitibmwatsonailab.mit.edu>

⁷ <https://research.samsung.com/artificial-intelligence>

Facial translations of still portraits are possible due to research “*Stargan v2: Diverse image synthesis for multiple domains*” by Choi, Y., et al, as their developed system model’s task is simple – to attempt to produce a good resulting “image-to-image translation” between different visual domains. In other words, domains that involve a collection of photos may be organised into a visually distinct category, with each image having a particular appearance known as a style. For example, we can create picture domains based on a person's gender, with the style encompassing cosmetics, beard, and hairdo. An ideal image-to-image translation system should be able to synthesise images in each domain while taking into account the various styles. (Choi, Y., et al, 2020)

Source: “Stargan v2: Diverse image synthesis for multiple domains” by Choi, Y., et al

Fig. 5. Excerpt of image synthesis of celebrities via domain translation system CelebA-HQ. (Choi, Y., et al, 2020)

Although this system only allows the synthesizing of singular images, it essentially replaces the manual Photoshop tasks significantly with impressive results. The research has also done a similar system network that applies to animal domain translations as well. Regardless, there are also other developing AI systems that take growing steps that seem more appropriate for films.

Recent AI research such as one from Stanford University named *"Deep Video Portraits"* by Kim, H. et al under deep learning "enables photo-realistic re-animation of portrait videos using only an input video. In contrast to existing approaches that are restricted to manipulations of facial expressions only, this system allowed the transfer of full 3D head position, head rotation, facial expression, eye gaze, and eye blinking from a source actor to a portrait video of a target actor." (Kim, H. et al, 2018) Entire regions of hair and the head, the upper body, and the environmental background behind the subject are all automatically rendered. The ability to create and modify video elements with head or bodily features is a challenge in computer-generated imagery as covered earlier when it comes to realism and believability, with "applications in video editing and movie post-production, visual effects, visual dubbing, virtual reality, and telepresence, among others." (Kim, H. et al, 2018) *"Few-shot Adversarial Learning of Realistic Neural Talking Head Models"* by Zakharov, E. et al brings this system type further by allowing moving head models on videos to be produced quickly using only a handful of reference data. The important feature of this suggests that it is capable to train the system quickly with only a few supporting images. However, at the research stage it unfortunately still requires a substantial amount of programming tweaking to the algorithms and parameters. Regardless, this design fundamentally allows the creation of a tool that can accurately learn detailed and realistic reference portraits of recorded people, as well as certain portrait paintings. More specifically, the "problem of synthesizing the photorealistic personalized head images are possible given a set of face landmarks, which drive the animation of the model." (Zakharov, E. et al, 2019)

There is usually great difficulty in rendering realistic computer-generated objects as well as good looking hair dynamics, but they are certainly achievable albeit with a considerable amount of effort and time required as covered earlier in this chapter. *“Neural volumes: Learning dynamic renderable volumes from images.”* by Lombardi, S., et al, have developed an AI neural model that can automatically model and render dynamic scenes and appropriate 3D reconstructed topology with just a few random shots from different angles around an object and does not require any form of extensive tracking to work.

Source: *“Neural volumes: Learning dynamic renderable volumes from images.”* by Lombardi, S., et al.

Fig. 6. Example of renderings of objects captured and modelled by the system. (Lombardi, S., et al, 2020)

Natural scenes generally involve complicated phenomena such as thin structures, developing topology, translucency, scattering, occlusion, and biological motion, making modelling, and rendering dynamic sceneries difficult. In these situations, mesh-based reconstruction and tracking frequently fail, while other techniques (such as light field video) often rely on confined viewing circumstances, limiting interactivity. (Lombardi, S., et al, 2020) In short, the model can predict volumetric mesh and representations and attempts at reconstructing the geometric shapes, as well as utilising ray tracing algorithms to render

translucent dynamics such as hair and smoke. This includes human physical instances as well, and the system can capture the complex facial structures as well as the fine details of human hairs in certain examples. The only limitations as stated in the paper is that the method does not consider refractive surfaces, and “high-frequency specular highlights are usually not correctly represented.” (Lombardi, S., et al, 2020)

In the recently published paper “*Adversarial Latent Autoencoders*” by Pidhorskyi, S., et al, further modified and trained generative adversarial network models under a new system architecture that goes by the name of ALAE, are capable of synthetically generating still images of an entirely fictional person to a high degree of accuracy and photorealism. (Pidhorskyi, S., et al, 2020) The demo code is currently released on the internet, allowing users to manually change and tweak facial features along sliders labelled, “male to female”, “smile”, “attractive”, “wavy-hair”, “young”, “big lips”, “big nose”, “chubby” and “glasses” while watching the system generate and morph faces in real-time. An application that will be able to, for example, construct synthesized fictional characters would perhaps be able to generate random background crowds and pedestrians in scenes when necessary.

Source: “Adversarial Latent Autoencoders” by Pidhorskyi, S., et al

Fig. 7. Examples of the ALAE system generating photorealistic portraits of fictional people.
(Pidhorskyi, S., et al, 2020)

The usability in image construction with a working user interface selection does put the system model into the feasibility perspective of visual effects, despite the limitation of resulting visible resolution that is usually capped due to the amount of hardware memory required to perform the tasks. The impressive feature of utilizing AI-assisted reconstruction of facial details while entirely omitting the requirement of a full computer-generated imagery preparation and production would introduce a wide new set of techniques that are invaluable to the visual effects industry as well as for many other applications around the world. In 2019, Japanese AI Company Data Grid ⁸ developed a system “that can automatically generate whole-body models of non-existent persons, identifying practical applications in the fashion and apparel industries.”
(Pangburn, D. J., 2019)

⁸ <https://datagrid.co.jp/en>

Allowing AI deep learning to recognize facial expressions and automatically map them from a reference actor onto a specified “target actor” image will allow the use of backup actors or stand-ins to act in shots where actors are unable to be present, providing a final digital render of the actor in motion that appears as natural and believable automatically within a short amount of time. Currently, this system can mean the reduced reliance on 3D computer-generated imagery capabilities for shots that require only the upper body or face, especially in conversation scenes in films. On average from some of the examples I have listed, the capabilities of deepfakes now are only producing small low-resolution results, which forces users to still resort to tracking & match moving to align these small rendered videos. The quality of natural and photorealistic movements is not perfect and can be easily observed and recognized by the naked eye thus far. These AI networks are still in evolving research and development, meaning deepfakes have not reached an acceptable quality to be considered for extensive visual effects production uses. However, should the process be perfected and released as a functioning system that can be integrated into the visual effects proprietary software of the industry, it will open up a wide range of possibilities in post-production alterations. Although numerous benefits would derive from such a technology, there are plenty of current issues regarding the premise of such “invisible” media modifications. In the mobile smartphone sectors, the popularity of Facetune and FaceApp demonstrate that despite the fears, consumers are still enamoured with the practical uses of deepfakes.

As this technology grows in its various ways of development such as applications on space exploration, automated voices that are synthesized digitally,

autonomous cars and the way visual effects can be produced for films, the general discussions and awareness “of identity theft, false information and digital confusion, would highly suggest deepfakes as a new distinct futuristic threat to democracy”. (Feinstein, L., 2019) deepfakes are currently also used for audio modified AI applications. Actor and director Jordan Peele demonstrated relative ease in producing a realistic-looking former United States President Barack Obama by utilising deepfake to recreate an accurate double of his speech patterns for a video. As such, there exists a significant potential of the new digital threat of illegal forgery that might be released to create misinformation in various political events such as the United States presidential election cycle or for any figure that will appear on recorded visual media. There is also a case where deepfakes can induce severe disruption from within the finance platform. Global cybersecurity company Symantec ⁹, “has reported several attacks on private financial organizations using AI. Three company’s ‘CEO’ called a trusted senior financial officer ordering a fraudulent urgent money transfer, imitating the CEOs’ speech patterns with an AI program that had most likely been trained on audio data taken from publicly available clips, many originating from social media posts. In these cases, millions of dollars were stolen from each company.” (Murphy, H., 2019) The highest-profile of misuse of deepfakes however still points to the dangers for women.

“Nonconsensual pornography accounts for 96 percent of deepfakes currently deployed on the Internet. Most target celebrities, but there is an

⁹ A brand of enterprise security software purchased by Broadcom Inc. in August 2019, <https://www.broadcom.com/products/cyber-security>

increasing number of reports of deepfakes being used to create fake revenge porn". (Adee, S., 2020) For example, an app that was taken down from the public named DeepNude used a similar AI system to "digitally undress" various women. By using the AI algorithm, one can simply upload an image of a fully clothed woman to digitally alter it to become photo-realistically nude. This demonstrates an example of deepfake pornography, a newly emerging online misuse of artificial forgery of real women faces onto pornographic bodies. Many women and actresses such as Scarlett Johansson were known to be a target on this controversial matter, and that the lack of control over such matters would mean that such techniques remain virtually lawless and unchecked. (Levine, N., 2019) These misuses do not just limit to still images or photos but expand to video related purposes just as easily. In addition, "women won't be the sole targets of such form of new bullying. deepfakes may well enable bullying more generally, whether in schools or workplaces, as anyone can place people into ridiculous, dangerous, or compromising scenarios." (Adee, S., 2020)

As such, there have been attempts to combat the potential issues stemming from the technology. For example, a former researcher of the University of Amsterdam Giorgio Patrini formed Deeptrace¹⁰ labs based in Amsterdam, Germany. The firm claims to serve "organizations by providing technology for monitoring and detection of deepfakes, and to date has uncovered more than 14,000 deepfaked videos." (Feinstein, L., 2019) Deeptrace is one of several new startups and academic groups that aim to build "deepfake detectors" to guard against the futuristic troublesome implications of AI-generated content.

¹⁰ <http://www.deeptracetech.com>

By utilizing generative adversarial networks models, it is essentially a measure of using AI to identify other AI altered media. “Adobe ¹¹ developed an AI-enabled tool that can spot deepfakes in digital images and has shared details on evolving software that lets users edit recordings of speech with the click of a mouse.” (Feinstein, L., 2019) It is highly likely that deepfakes will have a high degree of difficulty for prevention due to the technological potentials. Several United States laws regarding deepfakes have been implemented recently. “States are introducing bills to criminalise deepfake pornography and prohibit the use of deepfakes in the context of an election. Texas, Virginia, and California have criminalized deepfake porn, and in December 2019, the president signed the first federal law as part of the National Defense Authorization Act.” (Adee, S., 2020) However, many of the currently established laws only apply to areas where the illegal content creator resides in. When looking outside of the United States, South Korea, United Kingdoms and China seem to be the only other countries that are discussing actions against illegal deepfake cases. “The law commission from the United Kingdom is currently reviewing existing laws for revenge porn with an eye to address different ways of creating deepfakes. However, the European Union (EU) does not appear to see this as an imminent issue compared with other kinds of online misinformation.” There is “little evidence that the laws being put forward are enforceable or have the correct emphasis. While many research labs have developed novel ways to identify and detect manipulated videos such as incorporating watermarks or a blockchain” (Adee, S., 2020), it is still a difficult struggle.

¹¹ <https://www.adobe.com>

The “current state-of-the-art methods for deepfake detection can be easily bypassed if the adversary has complete or even partial knowledge of the detector. Therefore, there is a need for developing provably robust detectors that are evaluated under different attack scenarios and attacker capabilities.” (Neekhara, P., et al, 2020). The race between deepfake creators and detectors are, therefore, perceivable to be unending as the systems share similar constructions and would just rely on a repeated adaptation from one and another. More alarmingly, it is “possible to construct fake videos that are robust to image and video compression codecs, making them real-world threats since today’s videos shared over social media on the internet are usually compressed in the first place.” (Neekhara, P., et al, 2020). To help build their deepfake detector, social media sites such as Facebook have been working with researchers from various institutions who specialise in this field. “Twitter also made big changes to its policies, going one step further to plan in tagging any deepfakes that are not removed outright. YouTube reiterated in February 2020 that it will not allow deepfake videos related to any United States election, voting procedures, or the 2020 United States census.” (Adee, S., 2020) Regardless, this still leaves room for vulnerability for the rest of the world. The important prevention from the ongoing persistence of issues is currently at best, to create a growing awareness and educate the public on the possibilities of the technology, and to constantly fact check before passing judgements from possible fake news. It is clear that for visual effects creators in the industry, the technology itself proposes a promising new avenue of exercising more creative freedom and is stirring some excitement about the near future. The major worry is more on perfecting a believable result that will rival or complement existing workable solutions and the ease of autonomous

modifications with reliable controls to allow easy tweaking based on the production requirements. (As pointed out by interviews 2, 5 and 6 found in Appendix A)

Chapter 3. Automated Scene Manipulation

Many film scenes are a result of well thought out composites that blend live-action recorded footage alongside elements that have never existed and would be impossible to capture by the camera. Some processes of compositing also involve the post edits of cleaning up undesired objects or adjusting and utilising green screen in production sets from the live-action footage. In most scenes, the most notable elements are often the backgrounds and environments that are created to evoke a sense of realism to the world that the characters of a narrative would move through on-screen. Locations that have no favourable weather conditions on the day of the shoot, difficult and impossible hazards that will impose high risks for the actors and staff to be on-site, or just non-existent and fantastical all usually contribute to the need for some matte paintings, often resulting in what is commonly known as set extensions. Matte painting is a relatively old production technique that has existed from the early periods of film history. The method was “nothing more than pieces of black material, cut to shape and positioned in front of a camera in order to blank out part of a live action frame for later exposures with painted backgrounds.” (Edwards, G., 2015) This technique commonly allowed the combination of live action with painted environments – by means of double exposure of the painting with regions of blank space in the recorded scene and preventing the exposed portion of the frame from over exposure via masking. Another option was the commonly termed "glass shot", in which “the environment would be painted on site on a piece of glass positioned between the camera and the partial staged set.” (Edwards, G., 2015) Matte paintings have made significant achievements and created various iconic

moments from famous films in history, especially those done before the digital age such as “*Emerald City* in *Wizard of Oz* (1939), *Scarlet O’Hara’s* house in *Gone With the Wind* (1939), the giant warehouse in *Citizen Kane* (1941), to the futuristic visions of *Blade Runner* (1982) and *Star Wars* (1977) and *Indiana Jones* and the *Temple of Doom* (1984).” (Walker, J.)

As matte painters continued to develop the craft, they were constantly trying out new techniques to create environments that looked more photoreal than paintings. During the 1980s, “traditional photochemical effects techniques were being stretched to the limit... for many of the spectacular wide shots in *Star Wars: Return of the Jedi* (1983), multiple sections of live-action were rear-projected into gaps left deliberately empty in gigantic glass paintings.” (Edwards, G., 2015) Matte painting is currently still widely used in today’s film compositing techniques even though there have been various changes since adopted into visual effects. A scene is usually heavily edited and re-adjusted continuously as artists work closely with the directors and supervisors. As a result, to enable the shot to be changed quickly, digital technologies have enabled more control over creating several versions of set extensions instead of limiting to a single pre-planned artwork. Thus, the modern digital matte painting (DMP) is often created with virtual 3D environments that are filled with digital geometry that are mapped with manually created digital artworks or textures within certain software applications. Such digital matte painting environments would usually be made up of relatively simple geometry that is carefully placed based on the camera’s perspective on screen. The emphasis in such technique focuses on detailed layered textures that form the artwork onto these surfaces. It is not

uncommon for such textures used in digital matte painting to be made from actual reference images taken from real-world locations, or digital art paintings created specifically for the shot. The digital matte artist “refines these with retouching and hand-painting techniques, using software such as Adobe Photoshop or Foundry's MARI to create files that can subsequently be digitally projected around the necessary geometry, resulting in a rich, dimensional environment that can be viewed from several key angles or, if necessary, from all sides.” (Edwards, G., 2015)

There were also exceptional choices such as “*The Lord of the Rings trilogy* (2001–2003) that diverged from this tendency emphasizing instead the natural settings and utilising practical effects to achieve the result desired - a complex mix of actual locations, practical sets, miniature models, and digital set extensions” (Keil, C. Et al, 2016, p. 174–175). A relatively vital aspect of a successful composite of a scene generally comes from being able to match the colours between the matte painting with live action seamlessly. In today’s digital applications, one can simply retrieve the colour data from the desired film composition and directly transfer it to a digital set extension. Furthermore, “by merging and dissolving pixels, the blend line between live-action and painting can be made invisible” (Prince, S., 2011, p. 165). In creating dramatic visual aesthetics in scenes, lighting is essential. Today’s digital visual effects allow artists to change the lighting of a scene to various desired results in post-production to direct the audiences’ attention and mood of the screen composition. It is known and explained in recorded research on how and where “the eyes are attracted and will go first in a given composition” (Williams, D. E., 2006, p. 84) To present an overall more

pleasant and directed screen for the audience, post-production artists control the eye focus by adjusting contrast and modifying the lighting present in the film composites. More importantly, “digital lighting can be used to make the blending of an actor with the surrounding scene that is either created with miniature models or matte paintings or with computer-generated images to be more believable and convincing.” (Patruno, A., 2017) For example, actress Naomi Watts was lit up digitally during post-production from *King Kong* (2005) as “the lighting on [her] failed to match what was subsequently created for the CG (computer-generated) elements” (Prince, S., 2011, p. 64). In visual effects post-production, it is vital that compositing a scene looks correct and includes a significant aspect of photorealism to fully immerse the narrative for the audience. Looking forward, AI technological researchers are developing numerous deep learning systems that will prove invaluable to the digital set extensions in scene compositing.

For example, in the paper “Unsupervised Image-to-Image translation networks” (UNIT) by Nvidia, the AI research “aims at learning a joint distribution of images in different domains.” A domain is a classification of pixel characteristics of a species, gender, scene or object. For example, a group of images that contain images of dogs (of one domain) with another group containing images of cats (of a different domain). No hard pre-programmed knowledge or technique that is taught on directly translating an image from a group to another exists, such as making a dog look like a cat while retaining relatable fur colour and facial features. Through algorithmic pixel assumption techniques – a process of looking at particular pixels of a region on a screen and subsequently allowing the system to predict and estimate consecutive pixels

located near them based on trained data, images from such different categorised domain groups can be directly translated based on similar regions of pixel location on a screen. The system presents “high-quality image translation results on various challenging unsupervised image translation tasks, including street scene image translation, animal image translation, and face image translation.” (Liu, M. et al, 2017) This generative system model has since been heavily in constant reiterations and modifications, or even borrowed by various efforts and teams and is still undergoing more developments. For example, in the research paper “*Semantic Photo Manipulation with a Generative Image Prior*” by Bau, D., et al, their AI system can “accurately reconstruct an image and synthesize new content, consistent with the appearance of the overall scene on the image.” They created a system that is interactable by a user, allowing various semantic screen image editing processes such as changing an object’s appearance, removing objects automatically, or even digitally creating new objects that are fully blended with the base environment from within the scene automatically. These are processes that are already commonly practised in applications such as Adobe Photoshop for image editing, albeit in many more individual steps and clean up required. Given an application that can fully understand a given image or video and predict a relatively well blended and correctly positioned feature should artists choose to edit specific regions, it will provide digital matte painters or even compositors a great degree of freedom to include fast and reliable cover-ups to a final look for a scene.

However due to the model training, it is limited to environmental structures with a still significant degree of incorrect edits, namely on people,

certain bedroom decorations and vehicles as the system is unable to categorise them well enough yet.

Source: “Semantic Photo Manipulation with a Generative Image Prior” by Bau, D., et al.

Fig. 8. Example of the generative model that synthesizes new content that follows both the user’s intention and the natural image statistics. Top: Given simple user strokes, it can automatically synthesize and manipulate different objects while adjusting the surrounding context to match. Bottom: Users can edit the visual appearance of objects directly, such as changing the appearance of rooftops or trees. (Bau, D., et al, 2020)

The output resolution of the edited result is still limited to a small size and are not ready for a full high-resolution single image, let alone video for films. Optimisation of the editing regions take around 30 seconds each and may introduce various unwanted artefacts faster and lesser time allocated for the system. The system is also facing certain difficulties in reading user input operations – namely determining the vector of the drawn region in the space. As a result, the addition or removal of certain elements may sometimes cause various unwanted artefacts. For example, if a user wanted to remove certain chairs from a scene, there is a chance that the “rendered results do not cleanly remove all parts of the chairs,

and distorted parts that resemble chair legs remain.” (Bau, D., et al, 2020) In foresight, the system might have limitations in offering artists a more open-ended solution to changing scene objects or regions especially if the requirements were stylised and fantastical. Regardless, it is a notable step that attempts to incorporate user interaction and application into an automated scene generator based on few required sources that will prove very useful for pre-visualisation or quick renders of scenes during composition in post-production.

In the research paper “*High-Resolution Daytime Translation Without Domain Labels*” by Anokhin, I., et al, they “consider the task of generating daytime time-lapse videos and pose it as an image-to-image translation problem”. (Anokhin, I., et al, 2020) The system “combines a generative image-to-image model and a new upsampling scheme that allows users to apply image translation at high resolution”, notably without restricting to specified reference labels at all. It is essentially a model that can recognise the daylight from within given images, and thus output a result that is an automated time-lapse event predicted off them. It has also demonstrated some feasibility on paintings as well, basing off imitated styles.

So far, the experiments done have been on still images alone and have not developed into video output. The system currently does not support both translations of high-resolution images as well as an upsampling enhancement as a single working application, but instead, run them separately one after another. There are also no direct user interactions implemented. However, the ability to directly convert a scene’s time of the day will allow the ease of changing a scene’s mood and aesthetic considerations as a narrative direction, or simply a cost-

saving solution for when weather or location problems may arise during shooting. In the research paper *“Manipulating Attributes of Natural Scenes via Hallucination”* by Karacan, L., et al, they designed a two-stage framework that allows users to control high-level features of a natural scene directly. For example, the sky will have a reddish hue when the sun gradually nears the horizon during sunset. During summer, the grass gives off a more vibrant green that has some bright yellow hues. In autumn, the trees appear more brown and yellow.

Source: “High-Resolution Daytime Translation Without Domain Labels” by Anokhin, I., et al.

Fig. 9. Daytime translation results. Left – original images, right – translated and enhanced images (one style per column). (Anokhin, I., et al, 2020)

As a result, an environment will often be subjected to various visual changes according to various weather, seasonal or time conditions. “Such high-level changes are referred to as transient scene attributes – e.g. cloudy, foggy, night, sunset, winter, summer, to name a few.” (Laffont, P. Y., et al, 2014) The system can essentially translate scenes into various desired conditions such as

“different season (e.g. during winter), weather condition (e.g. in a cloudy day) or time of the day (e.g. at sunset).”

Source: “Manipulating Attributes of Natural Scenes via Hallucination” by Karacan, L., et al.

Fig. 10. Given an image, the system can generate different versions of the same scene in a wide range of conditions, e.g. night, sunset, winter, spring, rain, fog or even a combination of those. First, it uses a generator network to imagine the scene to its semantic layout with the desired set of attributes and proceeds to transfer the learnt style characteristics to the output without any reference style images. (Karacan, L., et al)

This outstanding system can generate relatively photo-realistic results as it can preserve the basic semantic details found in the images while transferring the newly targeted attributes of such desired conditions. The ability to consistently achieve this desired visual result is usually considered to be more challenging overall. For example, “for predicting the look of a scene at sunset, visual appearances of the sky and the ground undergo changes differently, the sky gets different shades of red while the dominant colour of the ground becomes much darker and texture details get lost.” (Karacan, L., et al, 2018). Most impressive is the system allows the direct input of user drawn labels, allowing a

high degree of flexibility to add or remove landscape features such as including a mountain along the skyline or changing a lake to a grassy field. (Karacan, L., et al, 2018). There are, however, some incorrect outputs as a result of the failure of transferring the desired style or simply errors produced by the visual quality of synthesized references. As a result, some of the elements that are translated are not represented with adequate photorealism with added artefacts created by the wrong blending of styles. Similar to the paper *“High-Resolution Daytime Translation Without Domain Labels”* by Anokhin, I., et al, the system also does not work with the integration of generator models and instead functions on two separate deep networks. Should the system work as an application with a reasonable amount of accuracy in its translations, digital matte painters and compositors will be able to quickly adjust a scene to the appropriate scene weather and season with control over the objects and elements present by selection. Further implementation of this into the visual effects workflows will allow production schedules to be planned even if the desired locale conditions and the season is incorrect.

Other than the excellent plausible uses of automated rotoscoping as covered in Chapter 1 in this thesis, rotoscoping unwanted elements from a scene of live-action footage often prove to be rather laborious. Some of the higher degree of difficulty in removing such unwanted objects such as fences or reflections without destroying the primary look of the shot is close to impossible without the backup of a clean reference of the similar shot, or entirely recreating with computer-generated imagery. In the research paper *“Learning to See Through Obstructions”* by Liu, Y. L., et al, the system allows for the removal of

“unwanted obstructions, such as window reflections, fence occlusions or raindrops, from a short sequence of images captured by a moving camera.” The system utilizes “the motion differences between the background and the obstructing elements to recover both layers” (Liu, Y. L., et al, 2020) via a deep neural network, recovering and reconstructing regions through warping and alignment resulting in data that are generated synthetically and applied onto real imagery outputs.

Source: “Learning to See Through Obstructions” by Liu, Y. L., et al.

Fig. 11. Example of recovery from a sequence of images taken by a moving camera through obstructing elements such as (a) windows, (b) fence, or (c) raindrop. (Liu, Y. L., et al, 2020)

The system will be able to separate both the foreground obstruction as well as the background into two individual layers. Integrating optical flow ¹ to provide more detailed autonomous estimation allows the system to generally “recover the underlying clean image from various challenging real-world sequences.” (Liu, Y. L., et al, 2020) As a result, the system has proven to work on traditional troublesome and laborious tasks such as removing some foreground obstructions such as a fence or some raindrops. However, errors in distinguishing

¹ Optical Flow, can be defined as “the motion of objects between consecutive frames of a video sequence, as a consequence of relative motion between the object and camera. We can say that optical flow is the distribution of apparent velocities of movement of brightness patterns in an image that arises from the relative motion of objects and the viewer.” (Sharma, R., 2020) It is a per pixel prediction and the main idea is that it assumes a brightness constancy, meaning it tries to estimate how the pixels brightness moves across the screen over time. (Gituma, M., 2019)

correct reflections on certain window conditions still produce incorrect separation of the window and what is in front of the reflections that resemble closely in nearby structures or colour regions. There are also no user interactions, which does not allow artists to have fine control over the elements in the scene which may lead to more work to clean up errors should they appear.

In looking at these few examples of direct scene manipulation by AI, it is clear that we are looking at some of the few instances of the beginnings of AI-assisted application tools that will eventually be available in current proprietary software of visual effects in the next few years. There is a distinct possibility of an entire look and development section in the visual effects post-production that will be covered and cleaned by such system models in the future, allowing artists to allocate more time for creative exploration and quick edits of not only set extensions but on the overall composited scene with a few minor tweaks and adjustments without having to make individual composition layer changes or re-renders.

Chapter 4. Artificial Intelligence in Digital Visual Effects

The visual effects industry has developed a systematic and robust pipeline that caters to film post-production efficiently over decades since the availability of affordable digital hardware and proprietary software back in the late 90s. As such, it is not rare to watch a film and notice a collective of visual effects studios providing work for a single major film, as the outsourcing of effects shots to various studios would mean that each studio can develop the best possible iteration of each shot with their best potential and within sufficient duration. As films come close to developing within a year or two alone in today's production expectations, so does the post-production budget and timeline that will inevitably always be pressured to deliver within close timings. The visual effects pipeline is no longer just a subsidiary – although there is the presence of the “fix it in post” mentality that is still present, the majority of the visual effects pipeline is now integrated alongside the main workflow of film production as the crew, directors, producers, as well as the supervisors all must have a collaborative direction through understanding and production standards. As a result, the visual effects pipeline involves the staging of planning, research, production as well as post-production as demonstrated in fig.1.

In every film production, there are three phases. The first would be the pre-production – the process and collection of work that is usually done before the actual filming production begins. This phase will take “a few weeks to over a year depending on the film.” (Whitehurst, A.) The second phase is the production which involves the actual filming. During this period, a lot of tasks are performed

by the visual effects team on-set simultaneously with work done in the company's studio. It is the visual effects crew's responsibility to ensure they have sufficient resources needed to build the intended effects based on the film's requirements, and these are a result of intensive research and pre-planning done to the best of the visual effects capabilities which might eventually fall short of final requirements unless prepared. These filming periods would usually take a couple of months ranging from three months to half a year. The third and final phase that follows is post-production. This period is when most of the final visual effects production functions including multiple contacting and liaising, resulting in multiple iterations and stages of development to results. During this time, the director's and lead editor's direct inputs will influence the film edits, music, and sound design as well. Research and development (R&D) are vital for projects since there is a need for proprietary software tools to be programmed or modified to allow efficient production solutions to new problems that have never been addressed before. As covered at the beginning of this paper, a general trend for visual effects is to push the limitations and capabilities of available technology, and the general audience often craves more and bigger effects every year. It would be close to impossible for an off the shelf industry-wide available software such as Autodesk Maya or Nuke by Foundry to fully encompass every production element as well as the inspired effects, thus research and development departments that consist of specialised mathematicians, scientists as well as programmers need to discover and attempt to modify or re-create programmable systems to fully realise those new tools that have yet to exist. It is usually more common for such tools to be made into plug-ins for existing proprietary software

such as Autodesk Maya. These tools can even be used across the entire visual effects company for future film productions.

In pre-production, a vital part performed by a visual effects company during this period is convincing clients that they have the necessary skills to produce certain effects. As such, numerous tests are done to demonstrate “a potential look, style or piece of technology that they may want to use in the production of actual sequences on the film.” (Whitehurst, A.) These are often proof that wins client confidence in showing off the capabilities of achieving possible results. As such, various highly experienced artists are usually assigned to deal with intense experimentation that is usually complex. When the proposed works and demonstrations win approval, the process of building the models can begin. 3D model assets are usually created in various qualities - low quality for pre-visualisation, medium that is normally used in animation and very high-resolution models for final renders. Often, such low-resolution models are used in tests and pre-visualisation (pre-vis). The 3D assets will therefore be produced as early as in pre-production. This pre-visualisation is a process of taking the storyboards or script descriptions of scenes and attempt to preview the entire flow of the movement and shots with the assets. The idea is to translate the storyboard into video imagery to visualise the timing of action sequences for camera movements and stage set-ups before actually spending cost for the production. The pre-visualisation is often made as a low-quality output. There will usually be multiple iterations of the pre-visualisation being showcased until an agreed final sequence is agreed upon. However, the pre-visualisations serve only as a guideline for film production. As most shots would be similar, some would

differ entirely during the actual filming as well. These are done so that departments and teams alongside the director will be able to see a preview and thus plan on how shots would look or otherwise function. On the other hand, there is also a possibility that a visual effects company would be provided with a substantial amount of specific supporting materials such as models and drawings when a client wants certain computer-generated imagery. These will be adhered to and reproduced during the final versions of computer-generated imagery. The various 3D computer-generated models will continue production until the beginning of the post-production phase. Most visual effect companies will usually try to complete every required digital asset just before they start creating the final shots of the films.

During production, a visual effects representative will usually be on-site during the film shooting of a scene or shot that the company is working on. Usually, they are the key persons that would know the overall scope of the project such as the project supervisors. Apart from scouting, they would be likely offering advice on various stage set-ups required for visual effects as well as acquiring numerous references required. These references will be used for lighting, texture, modelling and occasionally off-site environment building such as matte paintings. Certain computer-generated imagery requires knowledge and assets taken from real-world locations to be created effectively. Actors, various props as well as sets are often also digitally scanned into 3D. "LIDAR is a laser-based 3D scanning device that can create highly detailed models of small objects, buildings or environments." (Whitehurst, A.) However, these scanned models are often filled with various amounts of unfiltered data; they may have many millions of polygons

that will cause issues. As a result, a modeller is still required to build the assets manually while using the scanned models as a reference. By doing this, it ensures that final digital results and renders of the computer-generated assets will usually be more accurate or similar as compared to the actual physically scanned objects. Other than the reference images taken on set, the visual effects representative will also capture high-dynamic-range (HDR) images at the location allowing the technical directors (TD) of the lighting department to use current digital image-based lighting tools. They usually come as double 180-degree fish-eye ¹ images of the environment. They will be captured at various shutter speeds to allow more details of highlights and shadows to be collected as input data. They will eventually be combined into a single large file that will be used by the lighting as well as look-development departments.

In post-production, the film editing process starts and the live recordings from the shoot will be used. As most films are shot in digital formats now, they can be transferred around seamlessly over the internet to the various visual effects houses and departments in the film post-production. "Grading is a process whereby colour and exposure changes are made to the frames. The purpose of a technical grade is to get all shots in a sequence to have the same overall look and feel so that there are no noticeable jumps in brightness or colour changes when the shots are viewed together as a sequence." (Whitehurst, A.) The director of photography and colourist of a Digital Intermediate (DI) suite oversee the film's

¹ Taken from the Merriam Webster dictionary, "Fish-eye images are produced by a wide-angle photographic lens that has a highly curved protruding front, that covers an angle of about 180 degrees, and that gives a circular image." It is usually meant to create a visual distortion that is commonly hemispherical for panoramic purposes. In Visual Effects, this can be mapped onto a sphere to which most software are able to use it as a reference for lighting and texturing purposes.

final looking results. This is to ensure that before the production of visual effects is started on the shots, they are of the same look to remain consistent. The final ready film video shots and frames that require effects are called “plates”, which visual effects artists work on as a base. These can be sent to the various visual effects studio houses globally should the director and client decide to outsource into a simultaneous production. As 3D models are completed, they need to be prepared further for movement. A vast amount of technical understanding of the movements of objects and creatures falls under the category of rigging. Creature rigging is “especially difficult as layers of bone, flesh and skin need to be simulated by the rigging process. Once a rig has been created it is tested by animators who try and break it, then a new, revised, rig is created based on the results of the testing phase. Rigging may well continue throughout the animation cycle as better and more refined rigs enable ever greater results from the animators.”

(Whitehurst, A.) Match moves are also required in the animation department. A tracking department will produce the camera track using tracking software such as 3DEqualizer, SynthEyes, Nuke or After Effects. Sometimes tracker markers will be stuck on specific locations that will appear in the frame to assist in the task. More often than not there is a required adjustment of the tracker tool to achieve a digital virtual camera movement that aligns perfectly with the on-set camera perspective. Various body tracking and match move results rely on the best camera track available. Any computer-generated character or object movement that is translated from on-set is known as a body track. It is commonly used if you require a computer-generated element to be applied onto a character, such as tails, wings or other body visual effects so that the digital represented effect will be perfectly lined up with the actor. Junior artists who aspire to be technical

directors or animators in the visual effects business usually begin their careers from the camera and match-moving departments as they are often tedious and require fewer creative outputs.

Animation is the process of taking the model assets complete with rigged elements and bringing them to life as a digital creation, covering across creatures, characters as well as vehicles. Medium resolution 3D models are often used instead of high-resolution ones to not overload the working computers that render them during animating. This is vital for animators that require fast and reliable software to create natural movements. Animation showcases and demos are often rendered as grey-shaded videos commonly termed “play blasts”. These are used to gather feedback and thus the animation can be re-adjusted when required. Within the same department, effects animation refers to movements required simulation for realistic physics, such as fluids, body dynamics as well as particles. A technical director will use digital virtual characters and environments to provide the basis for their simulations. These often create impressive realistic looking fires as well as water results, including exposed and jiggling movements of flesh on creatures. E.g. By using specific geometries inaccurate scenes, the effects technical directors can then emit fire or water fluid simulations from them while digitally masking the same objects out and only retaining the desired effects during renders. This would assist the composition of the footage with the effects further down the pipeline. As the 3D models are in the final versions, they require image textures known as maps for characteristic details that are usually fine or not sensible to model such as skin and hair. “Using the reference photographed on set a texture painter can create maps for diffuse colour, specular as well as the

displacement and bump maps. These maps are often extremely detailed, sometimes 8K in resolution if the digital camera is to get very close to the model.” (Whitehurst, A.) The stage when textures are carefully refined with materials known as shaders, and virtual lighting on the objects to achieve the desired aesthetic is the look development. Technical aspects such as roughness, reflectivity, as well as shininess are factors that are adjusted by a look development artist to achieve photo-realistic equivalent or shot requirements. However, the visual effects supervisor and the film director will determine the desired result for unique objects where no real physical equivalent reference exists for comparison. The research and development department will often assist the look development in creating certain unique materials as well. It is not uncommon that these developers will program their shaders as a solution.

When the prior processes have been approved, the final finished computer-generated imagery will be rendered under the lighting department by utilising supplied material shader and the captured high dynamic range (HDR) lighting maps. These artists can take liberties of adding additional virtual lightings in a scene for creativity and specific stylistic choices. These lighting artists will “render the computer-generated imagery as different passes and then do a basic composite of their results over the film plate to ensure their lighting enables the computer-generated object to feel like it is in the filmed environment.”

(Whitehurst, A.) When done, the computer-generated imagery is handed over to a compositor. It is not uncommon to require placing a digital asset or character behind some real-world foreground obstruction. E.g. A computer-generated creature is moving just behind a physical on-set box. To achieve this. a

rotoscoping artist will trace the character as a matte to allow the composition to be combined well. This is also the reason why some shots use the green screen technique as explained in Chapter 1 earlier in this dissertation so that it would be more efficient to use a chromatic keying method. However, rotoscoping is still essential where chromatic keying fails in generating high-quality mattes. As covered before, this work falls onto junior rotoscoping artists. Compositing is where all the production elements are combined to form the final film, followed by a film-wide colour grade. To ensure that any visual effects are seamlessly integrated into the shots, compositors require a vast knowledge of techniques and artistic expertise. The director and supervisors will then do the final checks before approval for the final composites. These will be sent to the client in the form of digital files, which will be for the film edits which will eventually be released online and physical discs or bundled into Digital Cinema Packages (DCP) for release in cinemas. The visual effects pipeline is a complex and integrated network of stages during the entirety of a film's production and may even differ from a visual effects facility to another. However, the workflow that persists today still adheres to the standard of production.

Visual effects facilities "are constantly in search of tools that would provide better efficiency or novel looks. Beyond this motivation, there are normally three major acceptance criteria for the adoption of new tools: it will reduce labour costs sufficiently to offset the development time and resources, it will integrate sufficiently with the existing pipeline, and the usual software development risks should be acceptable." (Sylwan, S.,2011) Industrial Light & Magic with Disney Research are discovering new technologies capable of dealing

with facial replacement autonomously. The research also aims to find a technique that can capture an actor's performance without extraneous constraints such as gear equipment and tracking devices to allow them to perform better with comfort. "Anything that removes the friction between the artist and his/her interaction with tools would be a great step forward, and it is finding the nuances through an artist's day that are redundant or more technical." (Kaufman, 2019) As the AI technologies such as autonomous rotoscoping and deepfakes are extensively being developed in both academic and industrial fields, the digital visual effects industry already recognizes the invaluable contributions and advantage of AI inclusion to the existing workflows. If so, then what about other complex tasks such as seamless computer-generated imagery integration and set extensions? It might be possible for the future visual effects pipeline and workflow to be entirely different from the one the current industry uses today. For a while, "many of the most labour intensive and repetitive tasks such as match move, tracking, rotoscoping, compositing and animation, were outsourced to cheaper foreign studios, but with the recent progress in AI deep learning, many of these tasks can be not only fully automated but also performed at no cost and extremely fast." (Pennington, A. P., 2019) It is therefore of great interest to identify and propose the various stages of development and implementation of AI into the visual effects pipeline.

The following graphical figures are presented as a visual reference to better explain my vision of the visual effects pipeline that will be potentially adopted by the industry in the coming years. As a rough estimate of the emerging new technologies in AI and reviewing the requirements of visual effects

techniques that may use such technologies readily with ease, this process should take around 2 to 3 years for each evolving stage of change. Therefore, a distinctly different pipeline with full AI integration will likely be possible by the end of the decade. I will be proposing three distinct stages of development and transition concerning this assumption, showing my prediction of the change of particular departments and workflow. Each figure illustration is broken up into three major sections as explained earlier - pre-production, production, as well as post-production respectively. The arrows illustrate the workflow from one department to another. The straight-lined boxes each represent an individual department or work process that is commonly used today, where the dotted lines boxes each represent my interpretations of an AI applied department or process that will be included in the entire workflow.

Fig. 12. An expanded VFX Pipeline demonstrating how likely AI rotoscoping and deepfakes can be used in the near future as the first stage.

The first stage of development is demonstrated in Fig.12. As deepfake AI technologies continue to advance in development enabling the capabilities of creating more believable results in the visual and audio spectrum, before the operationally ready state it is likely able to serve for quick and rough edits during the pre-production stage. With enough research and development, visual effects studios will be able to resort to quick facial and vocal translations from an actor to a body double. This helps provide a quicker and rather efficient pre-visualisation deliverable to clients and for the director and supervisors, while also allowing much more time and freedom to experiment on the shots. In AI autonomous rotoscoping, it will be used to assist rotoscoping artists in reducing the need to rotoscope the majority of the frames in shots after the footage is delivered. Bash roto will be readily available quickly for compositors to start layering the shots together with the elements, plates, and various other computer-generated features such as characters and environments synonymously as every department expand in development. Therefore, final output corrections have a larger window for reiterations and remakes depending on the supervisors and director's direction and approval. As soon as the AI autonomous rotoscoping network has been fully realized and developed in a way to be integrated into current proprietary software safely, the lesser the need for junior artists to focus on the mundane and tedious repetitive tasks, allowing them to hone more into the creative composition department earlier and hence gain more tactile experience in their career growth.

The second stage is as shown in Fig.13. As the majority of visual effects related films involve the various complex digital edits and manipulations to actors to

allow them to blend naturally as the intended characters of the film environments, the capacity to quickly transform actors into final quality character results will cause drastic impact to current visual effects pipeline for quicker previews and planning during pre-production. In the presence of current real-time 3D technologies that can create on set environmental changes such as the Unreal Engine ² 5 being used in virtual production tools, as well as better visualisation technology, it is highly possible to have simulated on set previews and tests directly visualised on-screen so directors and supervisors can pull vital information before actual production and post-production starts. Even by utilising audio-based deepfakes, the actors' accurately synthesized voices could be heard during replay despite recording only a body double. There will be a lesser need for tracking & match-moving tasks as deepfakes automatically apply modifications to the screen pixels regions of the actors. Modelling, rigging, texturing and animation of the usual computer-generated assets that would have been used on the tracking results to replace might also be less important, as deepfakes would also completely produce instantaneous photorealistic character results down to the skin details, movements and lighting conditions. Instead, those departments would most likely be focusing on computer-generated environments and fictional unnatural characters with unique features that AI is unable to fully comprehend. This perhaps suggests that the modelling department would still be required to work closely with the pre-production planning tests and pre-visualisations to cover the gaps in which the AI systems are still unable to correctly identify and produce. As the research and development of deepfakes

² <https://www.unrealengine.com>

allow full integration with existing proprietary software either as a plugin or a standalone program, the visual effects industry will expect to use it throughout film production. A visual effects industry-standard of deepfake should allow a simple click-to-transform solution while giving artists the capabilities to make distinct parameter changes to facial and bodily movements that are naturalistic in a non-destructive process³. Likewise, there should be capabilities to entirely swap character face and bodies to various other assets seamlessly while retaining the animated parameters. With those developed, likely, deepfakes data developed during production on set, or even newly constructed ones during post-production based on recorded footage are available readily for compositors to use. The results of deepfakes will be also highly likely compatible and developed with AI autonomous rotoscoping processes.

³ Non-destructive processes are workflows that allow the user to make changes to a work and still have the choice revert to the source without destroying the entire project or having the need to make additional changes.

Fig. 13. A revised VFX Pipeline demonstrating how likely AI Rotoscoping and deepfakes can be used in the near future during the second stage.

AI autonomous rotoscoping processes should eventually be able to perform optimally with little human input. The system should be able to distinctly segment various elements of objects and actors from backgrounds in real-time. This will open the various possibilities of live compositing and tweaking of potential background swaps on the fly so directors and supervisors will be able to fully visualize how the shots would likely appear before the final output. Such AI networks have already been in development and used in various mobile applications that allow users to be able to automatically capture their facial portraits or entire bodies, and simply swap the backgrounds with a flick of the index finger. It will, therefore, no longer be a distant idea that live real-time previews of AI-augmented footage will be able to do the same. AI autonomous rotoscoping will also be able to help eliminate the need for many junior rotoscope artists during the post-production phase, instead of being able to accomplish the laborious task quickly once the visual effects facility receives the footage and relying upon only a few rotoscope engineers and the supervisors to do minor tweaks and adjustments before the thousands of final quality frames are ready to be sent to the compositors for compositing.

A final fully integrated stage is demonstrated in Fig.14. when deepfake and autonomous rotoscoping as well as automated scene manipulating networks have been fully developed and optimised for close to instantaneous and hyperrealism per pixel results. We can see the significant change in requirement for the character replacement and creation techniques resulting in a reduction of primary visual effects production as well as post-production roles involved in the character modelling, rigging, animation and lighting & rendering departments.

These complex tasks require a lot of time and are expensive to generate satisfactory results, and the alternative solution that is not reliant on computer-generated capabilities would allow visual effects companies to generate faster composition of characters in scenes. There should still exist a modelling department in most visual effects facility, however, as fictional recreation of materials and creatures and sets will likely still be around, albeit with greater autonomous assistance with the AI integrated software that would be able to generate 3D mesh topology off referenced images conveniently. As a result, there would be more time for future edits and effects with lesser concern over how natural and photorealistic the characters appear on the screen. During some instances of set extensive asset creation, environmental creation via AI synthesised processes will be fully used as well.

Fig. 14. An AI integrated VFX Pipeline demonstrating how likely AI can be used in the near future during the third stage.

For example, footage that is shot on a clear day can be translated into a night or rainy shot with a certain believable degree. Certain landscape structures and natural formations can be selected by an artist to remove, add or even replace down to the pixels and blended appropriately and automatically with AI assistance. A normal face of an animal such as a cat can be transformed in hybrid computation and translation into a face of another animal such as a lion with unique features such as hair or a bone structure from a dog. However, automatic scene replacement, character facial, bodily features modification and crossovers are potential uses of unsupervised image-to-image translation networks that could allow visual effects companies to save cost and time in reshooting entire shots in different sets where factors such as weather and environmental issues would have caused delays and additional post-production costs, and to reduce the need for computer-generated creatures and character creations. Most of the systems I have covered in the previous chapters currently lack the quality to be readily used in visual effects production – a major concern by visual effects artists and directors as covered in Appendix A. If the system contains adjustable parameters with the ability to allow artists to modify scenes with customized design elements, this could allow the translation of characters and environmental scene conditions in shots to fully integrated in a naturalistic way. Tracking and match moving will no longer be a laborious requirement as AI tools will automatically integrate the various moving layers naturalistically. Working alongside technologies from autonomous rotoscoping, artists will be able to use final quality mattes from the beginning of production to modify and translate characters, creatures and elements of the environmental details and conditions that will be the background elements. Using deepfakes will allow the artists to do

actor facial or bodily corrections, replacements and expressions thereafter that are easily recognizable. This allows better training for the networks to allow sufficient adaptation and rendering to a believable quality. The training starts from pre-production tests and pre-visualisation to post-production development, maintain consistency and allowing the AI networks to fully understand and perform optimally for synthesizing processes during fixes or replacements. The look development, lighting and colour grading departments will likely require fewer people to adjust the settings creatively with AI automatically mapping and pixel correcting their respective on-screen conditions. It is through these possible combinations that could potentially revolutionise the capabilities and existing industry visual effects roles, generally bearing a reliance on AI to assist in the laborious and complex processes while allowing artists to give sufficient correction to details for a desired artistic direction. These applications of AI “solve a variety of problems for accelerating visual effects processes, enabling truly interactive character creation, and expanding pipelines for animation and simulation data.” (Failes, I., 2019)

After placing such estimations in applying reasonable AI usability into the respective departments of the visual effects production, the next phase of enquiry would likely be – how would the entire AI technology be adopted into the industry in the first place? There are various avenues of possible entry this technology gets adopted into the visual effects pipeline. As such, there are some possible scenarios that I believe should be considered in this respect. The first would be that big visual effects companies like Digital Domain and Industrial Light & Magic would still maintain their forefront legacy, and with their large monetary investments

that could be used in gaining a distinct advantage for in house research and development of AI systems for visual effects. A prominent recent example where AI has been used in visual effects is Thanos in *Avengers: Infinity War* (2018) by Digital Domain. Working alongside visual effects company Weta Digital, Digital Domain's "artists rigged the character model to provide controls for animators, created textures and shaders that determined how the purple skin would react to lighting conditions, and set up muscle simulations and cloth dynamics. They grew fine hair on Thanos's body and arms and stubble on his head, as well as gave him armour and a gauntlet. The team used Autodesk Maya for modelling, rigging and animation; Chaos Group's ⁴ V-Ray for rendering; SideFX's ⁵ Houdini for effects; a proprietary Maya plug-in called Atomic for lighting; and their custom Direct Drive for transferring motion capture data to the computer-generated character."

(Robertson, B., 2018) They then used a uniquely developed machine learning tool known as "Masquerade" to "transform actor Josh Brolin's face – captured with head-cam cameras looking onto facial tracking markers – into the film's leading villain character. This involved taking advantage of facial-capture training data."

(Failes, I., 2019) Various props and sets contained multiple cameras to capture motion data, and actors wore motion capture suits when performing. When certain computer-generated characters require a scene with dialogue, the actors wore "a helmet equipped with two high definition 60fps cameras pointed at a set of about 150 dots" (Robertson, B., 2018) that were placed on their faces. Any performances by the actors were captured real-time as motion data while they performed live on set. A low- resolution model of the character moves with the

⁴ <https://www.chaosgroup.com>

⁵ <https://www.sidefx.com>

motion data captured. This motion data is also adjusted on the fly to bring out as much subtle movements and details performed by the actors before applying them onto actual computer-generated models. In certain cases, actors also record a different capture for higher resolution data. For the case of Josh Brolin's data, the team from Digital Domain also resorted to capturing high-resolution facial data, to which "his expressions, the limits of his facial range, how his skin wrinkles, and so forth by using Disney Research's Medusa ⁶ system". (Robertson, B., 2018) They applied the captured data into Masquerade, to train and learn how Brolin's face moved in high detail. Meanwhile, they also transferred some of the low-resolution details captured from the helmet cameras. "Masquerade converted the 150 facial data points from the live performance into roughly 40,000 points of high-resolution facial motion data. The algorithms essentially converted the low-resolution data into high-resolution data based on the system's knowledge of Brolin's face." Darren Hendler from Digital Domain shared that the company "already knew that they could build a system that will take motion-capture data and produce a result" (Failes, I., 2019) with machine learning, and further feed corrections to the system as it advances. As a result, future results can continuously be corrected in the desired manner. Visual effects supervisor Kelly Port from Digital Domain explains that Masquerade was very capable of doing such high detailed conversions well, and the AI system would continuously improve the more data is learnt. The trained system became more accurate and

⁶ The "Medusa Facial Capture system, developed by Disney Research Studios, consists of a mobile rig of cameras and lights coupled with proprietary software that can reconstruct actor's faces in full motion, without using traditional motion-capture dots. The technology comes as the result of many years' worth of research and scientific advances in capturing and modeling of human faces. Medusa delivers high-resolution 3D faces, with the ability to track individual pores and wrinkles over time and can recover per-frame dynamic appearance (blood flow and shininess of the skin), providing a very realistic virtual face that is ideal for creating digital doubles for Visual Effects and computer games. Medusa is useful in two primary scenarios: 1 – building an expression shape library, and 2 – reconstructing performance dialog." (Medusa.)

could achieve intelligent results. Despite the need to manually place some minor adjustments, Masquerade saved much production time while ensuring the motions were sufficiently detailed. He mentioned in an interview that he thought of taking a low-resolution mesh capture from the live performance to contain the same fidelity as if it had been captured separately in high-resolution with Medusa as magical. (Robertson, B., 2018) It is therefore rather clear that AI-assisted tools will be able to benefit visual effects production in allowing cost-effectiveness and efficiency. By my estimation, the majority of the developing AI tools will be readily available in big visual effect houses during the next few years of 2021-2025. In a similar cycle that was seen during the digitisation period, the tools for in house development will only be taken out and be readily sold as a commercial suite after they have successfully led the industry into the next AI-assisted production workflows. When these AI systems slowly disseminate, only then will medium-sized commercial visual effects houses benefit when they start to adopt them into their workflows as an industry practice.

The second scenario will come when industry established software companies like Autodesk who are already affiliated with visual effects as a sole proprietary developer would successfully integrate their research and development with AI systems into the existing software that will be widely accepted in the visual effects industry. We have covered the aspect of rotoscoping AI by the means of developer Kognat on Rotobot earlier in Chapter 1, for example. As a result, any visual effects company that are already invested in the software for their projects will easily adapt and have an AI-assisted upgrade to their licenses. This will allow a simultaneous evolution of visual effects workflows in

the industry on a similar level or at the very least a concrete base where the industry can work with certain AI systems as a common fundamental level integration for production. Another example - match moving is usually a vital and complex task and can at times require a significant amount of time when certain manual corrections are required. Software developer Foundry has developed AI algorithms that is able to “accurately track camera movement using metadata from the camera at the point of acquisition such as the lens type, as well as how fast the camera is moving etc.” (Pennington, A. P.,2019) Alastair Barber who is the lead software engineer mentions that “the results have improved the match moving process by 20% and proved the concept by training the algorithm on data from DNEG (Double Negative)”, an established visual effects company. Developer Arraiy has developed a system which can place computer-generated imagery moving objects into moving scenes photo-realistically. For example, they “provide a tool in which an artist can train a neural network to create mattes for an entire shot by providing the networks with a few labels on what the artist desires to segment. Once that network is trained for that particular shot, the algorithm can generate mattes for an arbitrary number of frames, and even in real time, while on set, if desired.” (Failes, I., 2019) The Mill ⁷ is a facility that is developing AI that demonstrates real-time capabilities of commercial, film and television broadcasting productions. The Mill used Cyclops and Blackbird which was an adjustable rig attached to a car that is capable of capturing motion data from the environment with company Epic’s Unreal Game Engine for a collaboration project named ‘The Human Race’ project. During the shoot “Cyclops stitched 360-degree

⁷ <https://www.themill.com>

camera footage and transmitted this live to the Unreal Engine producing an augmented reality image of the virtual object tracked and composited into the scene using the AI computer vision system from Arraiy.” (‘The Human Race’) The result was a real time rendering of a car on an actual live footage capable of reacting to lighting and environmental changes where interactive users can change these conditions, and thus capable of customising majority of the features on the fly with photo-realistic visuals. Currently the technology is only being used in car showrooms and demonstrations, but one could wonder how it may contribute to visual effects in the near future. The capabilities to render or capture digital objects in real time with live action locations is but a start to realising the potentials of virtual production. Ziva Dynamics ⁸ has AI algorithms that “are trained on physics, anatomy and kinesiology data sets to simulate natural body movements including soft tissue movements like skin elasticity and layers of fat, resulting in the possible creation of computer-generated imagery characters and creatures at a fraction of the time and cost of traditional visual effects processes.” (Pennington, A. P., 2019) RADiCAL ⁹ “was founded to develop the world’s most powerful computer vision technology focused on detecting and reconstructing 3D human motion from 2D content.” (‘AI-powered 3D animation’) These systems are meant to provide an AI solution for motion data when no physical data is available. For example, one could upload a recording of an actor onto a cloud service where the AI system will translate the data into a motion data available for animated movements in computer-generation. They “are using multiple bodies of science in a proprietary configuration which includes computer

⁸ <https://zivadynamics.com>

⁹ <https://getrad.co>

vision, deep learning and anthropometry. The system continually learns and perpetually improves” (‘AI-powered 3D animation’), and “the latest version promises twenty times faster processing and a dramatic increase in the range of motion from athletic to combat.” (Pennington, A. P., 2019) Likewise, San Francisco based firm DeepMotion ¹⁰ “uses AI to re-target and post-process motion-capture data. Its cloud application, Neuron, allows developers to upload and train their own 3D characters — choosing from hundreds of interactive motions available via an online library.” Thus far, DeepMotion’s suite of tools also provide end-to-end solutions including single camera, AI-based body tracking and animation generation, Virtual Reality (VR) full body avatar tracking using minimal trackers, a powerful full articulation physics software development kit (SDK), and other breakthroughs in character interaction and digital motion. As such, the constant research and attempts to bridge AI into the visual effects industry has already begun, and I believe that this might be the most plausible scenario of the question in point.

A third scenario would be that specialists from fields outside of the visual effects industry that have experience in computer science etc., will decide to develop new tools aimed to entice the industry companies into bringing them into the field of filmmaking. These are similar to the various academic research we have covered thus far in the thesis, with researched and developed solutions catered specifically for visual effects. This entry has the potential for discovering revolutionary solutions that the visual effects industry might take interest in changing fundamental workflows. However, it is also more akin to further

¹⁰ <https://www.deepmotion.com>

redevelopments of the AI systems to better fit for specific effects required. Most likely, such research forms the basis of interests for independent developers to start bridging the algorithms with specific visual effects toolset assistance. CEO and co-founder of Pinscreen ¹¹, as well as researcher Hao Li, shares that since “neural networks have existed for some time, but it was only relatively recently that deep neural networks consisting of multiple layers in the system can be trained efficiently with GPUs and massive amounts of data. Deep learning-based techniques outperformed many, if not most, of the classic computer vision methods for fundamental pattern recognition-related problems, such as object recognition, segmentation and other inference tasks. Since many graphics-related challenges are directly connected to vision-related ones – such as motion capture, performance-driven 3D facial animation, 3D scanning and others – it has become obvious that many existing techniques would immediately benefit from deep learning-based techniques once sufficient training data can be obtained.” (Failes, I., 2019) As covered in the previous chapters, more recently, various research on AI have developed new systems that can synthetically generate believable scenes or objects from little to no data input. The field of computer vision relating to variational autoencoders and generative adversarial networks (GAN) is responsible for such developments. In this thesis, I have selected the generative adversarial networks that appeared to be relatively associated with the needs of visual effects, and thus I would assume that the tools would eventually adopt these varied technological techniques.

¹¹ <https://www.pinscreen.com>

The final scenario I propose would be that the AI systems inclusion might not even be heavily adopted for big visual effect companies since they can still retain quality production with the established visual effect workflows but instead is more beneficial for small or independent companies where labour and cost are of higher concern for their commercial projects. As such, perhaps independent developers who develop certain AI hardware or applications that are not related to visual effects in the first place might garner enough attention and utility for some commercial productions. As such, the smaller visual effects production houses will more likely adopt such tools in hopes of saving costs during production. China has a few notable small AI software incorporated into such small companies. An example would be Insta360, an action camera device developed by Insta360, a business of Arashi Vision Inc.¹² that incorporates AI systems built within as an update to their devices. Action cameras are very versatile and often come with accessories that allow them to be attached to wearable gears such as headwear and vests as they often come in small designs. Aimed purely towards consumers who seek affordable and convenient portable camera devices to capture almost any desired moment when the traditional camera set up would be unfeasible in the first place, the action camera is one of the new iterations of camera hardware engineered to shoot at high-quality videos with as much automated assistance as possible so that users may focus on their activities to the fullest. The Insta360 uses AI to allow consumers to get capture high detailed recordings from the cameras. PureShot is the new update where AI is used “to enhance the dynamic range of photos while reducing noise and

¹² <https://www.insta360.com>

preserving detail, resulting in a vibrant image recorded. What used to take multiple shots combined into one with auto exposure bracketing, now just takes a tap with PureShot.” AquaVision is another feature included for consumers shooting underwater where the system is capable of balancing and correcting underwater colours automatically when there is a lack of red hues. The result “is vibrant and clear underwater video renders, without the need for external filters or laborious editing.” (Insta360, 2020) Such action cameras may very well provide sufficient reason for investments for various forms of film production requirements at a much cheaper alternative and perhaps even challenge new technical aspects of film making when former shots and visual effects may be unable to do due to the camera rig setups etc. Likewise, in various other commercial applications, AI is currently an important evolving technology for smart learning interfaces. Holographic Interfaces ¹³ - “a boutique research, design & development shop specializing in Augmented and Mixed Reality experiences”, and its co-founder Tyler Lindell who is also an Augmented Reality (AR) focussed software engineer assisting SpaceX's ¹⁴ Hyperloop partners, explains “One might go so far as to suggest that augmented reality relies on AI to be effective. AI could be used in augmented reality as well as virtual reality to predict the interface a user might want or need within a given situation and show options for a User Interface (UI) - or automatically bring up a UI that is perfectly suited and timed.” Augmented reality “can deploy AI for object recognition and tracking - as well as gestural input. Gestural recognition and tracking are, what allow people to use their hands to manipulate 2D and 3D objects within a virtual space using specific

¹³ <http://www.holographicinterfaces.com>

¹⁴ <https://www.spacex.com>

movements. The incorporation of eye-tracking and voice commands as a means of manipulating the virtual environment are equally relevant.” (Daly, C., 2017) These systems form a basis for various smartphone apps and apparel devices where camera footage is capable of being altered live. For example, there exist experimented projects in the fashion and apparel industry based on “smart mirrors”. These mirrors can augment a person’s reflective image, adjusting and tracking virtual clothes, hair properties, as well as facial, make-up in real-time realistically. L’Oréal ¹⁵ released “an app called Style My Hair in partnership with ModiFace” allowing user previews by adjusting and changing the look of their hair digitally live. The system produces realistic detailed renders on individual hair strands and can be easily changed to other varied selection by a swipe of the hand. It is very helpful when consumers can virtually visualise themselves before committing to an actual hair change. Perfect Corp. ¹⁶ developed the YouCam Makeup-Magic Selfie Cam which essentially does the same with additional virtual features of changing eye and makeup colours. Shopping for cosmetics usually involve taking makeup on and off whilst trying different “shades of highlighter, foundation, blush, lipstick, and eye shadow. Being able to do this virtually prevents any adverse reactions, especially for sensitive skin and makes experimentation easier.” (Marr, B., 2019) In the clothing industry, Zara ¹⁷ experimented with its augmented reality displays in conjunction with their Zara AR app for smartphones. When customers point their smartphone cameras onto a store window or certain sections within, they would be able to preview models

¹⁵ <https://www.loreal.com/en>

¹⁶ <https://www.perfectcorp.com>

¹⁷ <https://www.zara.com>

wearing the clothes augmented onto the location, allowing a virtual assessment before purchases. The experience presented augmented reality “images of international models - including Lea Julian and Fran Summers - posing, strutting, and speaking on user's phone screens for short blasts of 7-12 seconds. Customers were able to view 12 different scenes, all of which were captured with 68 cameras on a 170 square meter stage.” (“Zara's Augmented Reality App Brings Virtual Models to Life in Stores”) Augmented reality developer Fitnect¹⁸ used Microsoft’s Kinect AI software to develop an augmented reality fitting room. The application “features depth and infrared sensors that reside in Azure, Microsoft’s AI Cloud service.” The Kinect was formerly developed as an Xbox gaming technology, and the recent versions come with AI-assisted applications that enable depth reading that allow Fitnect to work with objects that are captured at various distances in a room to be rendered onto the screen accurately. Fitnect “is connected to a database of 3D images of clothes and accessories hosted on the Azure remote cloud servers. Shoppers can search and select clothes and accessories from a retailer’s catalogue with various hand gestures. They can also gesture in the direction of a specific item to it virtually. The system also features real-time cloth physics, which the company claims can show what the clothing item looks like when a user moves during full-body motion capture.” (Jesus, A., 2019) Although such cases may seem to differ to an extent from typical visual effects applications, the capabilities of editing video footage in real-time are a key potential crossover. It essentially covers similar fields of replacing or adjusting on-screen character parameters and aesthetics. The more the smaller companies can develop

¹⁸ <https://www.fitnect.com>

innovative and robust yet reliable devices and applications, the more variation of tools that will likely be a possible adoption as a benefit by the visual effects industry as a whole.

Chapter 5. Other Technologies for Visual Effects

An important investigation under this chapter which will be covered later revolves around another prominent and adopted production process that is made possible with the latest technological advancements – virtual production (VP), which is set to change the way the typical visual effects pipeline work apart from AI. Before that, there are a few more notable instances of technology that I believe would be very useful if they are eventually adopted into visual effects utilisation as well, which I have not included in the prior sections of the previous chapters as they appear in their unique respective fields. Everything covered in this chapter still facilitates potential improvements to the current workflow of the visual effects industry.

Source: “learning predict-and-simulate policies from unorganized human motion data” by Park, S., et al.

Fig. 15. Example of the predict-and-simulate system allowing agile, interactively controllable, physically-simulated character to be equipped with various motor skills learned from unorganized motion data. (Park, S., et al, 2019)

For example, the research on realistic movement physics “*learning predict-and-simulate policies from unorganized human motion data*” by Park, S., et al, “allows physically simulated biped characters to be equipped with a rich repertoire of motor skills... Creating a virtual character from a collection of human motion data

has been a long-standing problem in computer graphics... It is challenging to understand what actions are available from unorganized motion data sets, how the actions generalize, and how to generate a sequence of plausible actions to animate a virtual character.” (Park, S., et al, 2019) As a result, the widely used assets of crowd simulations in visual effects are usually pre-programmed and in certain instances still rely on prior animation or motion capture techniques. By adjusting the objective parameters, the user can simulate different ways characters will behave and move, reacting in real-time to environments and other character bodies. Should there be direct incorporation of random simulated physics from a set of input conditions and data, background crowd simulations for visual effects will be much easier to produce in believability and detail.

Source: “Learning temporal coherence via self-supervision for GAN-based video generation.” by Chu, M., et al.

Fig. 16. Example showing off the high-quality video super-resolution feature of the AI system. (Chu, M., et al, 2020)

Super-resolution enhancements will be available thanks to system research such as “*Learning temporal coherence via self-supervision for GAN-based video generation.*” by Chu, M., et al, that makes use of a generative adversarial network function that can be trained, predict and render coherent details on low-resolution images. The system is currently still unable to work on full working videos as of now, but only on still unmoving images. Should the process be developed further with video features, such technology will be able to recover any significant footage or texture losses during the visual effects production, which may prove convenient and helpful for artists to work with less concern and limitations.

Source: “Portrait Shadow Manipulation.” by Zhang, X. C., et al.

Fig. 17. Example showing off the removal of shadows from portraits. (Zhang, X. C., et al, 2020.)

Similar to this field of video data recovery or manipulation, “*Portrait Shadow Manipulation.*” by Zhang, X. C., et al, allows users to be “re-lit” in post-production in a controllable way. By controlling various lighting properties, it “can alter the mood of an image, direct the viewer’s attention, or tell a specific story.” (Zhang, X. C., et al, 2020.) Aesthetic qualities such as the position and

softness of shadows, as well as the lighting ratio between the bright and dark regions of the face, are usually dictated by environmental restrictions. (Bau, D., 2020) To address this issue, it is a common practice during film production to use light bouncing cards and flashes. Sometimes there are requirements to do computer-generated replacements and relighting in post-production correction as well. In this research, the AI system is capable of automatically generating photo realistic light balances, remove or reduce unwanted dark unlit areas and even help feather harsh shadows. Unfortunately, the research is also only starting on single image properties, and not for video data presently. The premise and future impact of this capability is something I presume to be of large assistance to visual effects artists soon.

Of recent developments and interest in visual effects production processes is that pre-production is getting more heavily invested in preparation for a more believable and efficient production during filming itself, with the use of full LED screen setups for actual shoots that allow seamless light and background integration onto live subjects on camera for added realism and ease of film-making. As a modern form of the rear-screen projection method of in-camera techniques that derived back from the 1960s, current virtual productions can produce correct parallax to the camera movements, while providing high-quality real-time assets that encompass the actors on set. This form of virtual on-set production has already proven to work in a few films with impressive results. *Oblivion* (2013) is a film with visual effects by Pixomondo ¹ that made use of a massive projection technique by using recorded video plates that are merged in

¹ <https://www.pixomondo.com>

Foundry's Nuke software, with computer-generated elements creating a 20-minute clip with the dimensions of approximately 13000 pixels wide and 2000 pixels high collection of images. The projection was 152-meter-wide by 12-meter-high with 21 projectors on set, creating an environment in which the actors were able to immerse in, as well as contain realistic lighting and reflections required for the in-camera shooting. (Seymour, M., 2020) *Gravity* (2013) is an Oscar-winning science fiction film that emerged as a box office success of around USD 723.2 million. It has used such virtual setups effectively during film production as well. Visual effects house Framestore ² made a cube that was used like a lightbox which was hollow inside. It was around 3-meter-wide and as high as 6-meter-tall. Inside on each wall, there were 196 panels each containing 4,096 LED lights and measured 60 centimetres by 60 centimetres individually. The stage was a setup of rails with an ARRI ³ Alexa camera fixed on a robotic arm supporting a Bot & Dolly ⁴ IRIS rig. The lightbox cube was adapted from technology deployed commonly during that period, such as creating large screen visuals as the backdrop of music bands or used in big sports stadiums. Every pixel is an LED that could be controlled, so controlling realistic affecting lighting on the actors were purely digital instead of physical. As a result, controlling the environment around the actors could be done digitally at any time without the need of moving the screens physically.

Recently, the virtual production stage workflow-enabled complicated shots with visual effects to be captured using game engine technology that is real-

² <https://www.framestore.com>

³ <https://www.arri.com/en>

⁴ <https://www.botndolly.com>

time and in-camera, with the combination of surrounding LED screen technology. Photo-realistic digital landscapes and environments can thus be used live, greatly reducing the requirement of traditional green screen methods in visual effects, allowing a kind of real-life working simulation style of production. This entire production process works since the background automatically adjusts the parallax to the physical recording camera's perspective – the stage set-up works due to the LED stage directly matching up with the motion capture system data. *The Mandalorian* Season 1 (2019) used such a production technique and encompasses over fifty percent of the workflow. This resulted in a less significant requirement of location shooting as the environments were virtually set up using the stage. As a result, all the actors in *The Mandalorian* performed in a big virtual stage of a 270-degree semi-circular wall of 6-meter-high by a ceiling of 22-meter diameter wide. (Seymour, M., 2020) The set pieces were a mix of digital as well as practical. Digital environments were controlled virtually on the walls and adjusted on the fly during shooting. This allowed the NVIDIA Graphics Processing Units (GPUs) to run accurate tracking and perspective correction renders at very high resolutions. The entire virtual production prompted a new change in the visual effects workflow - final creative directions were made at a very early stage. This was real-time live compositing with a game engine used to produce a finished quality look during filming. During periods when issues were filming certain shots using virtual production, an Industrial Light and Magic team manually changed and replaced sections of the wall with traditional computer-generated 3D. However, this method was generally cleaner than traditional green screen techniques due to less green spillage and lighting restrictions on actors. When compared to a typical green screen set up, the virtual production walls

allowed realistic reflections and highlights on characters that wore reflective suits. This significantly reduced the time required to track actors with such surfaces and recreated the desired surface details using traditional digital 3D methods. This entire virtual production workflow technology was headed by Profile Studios ⁵, Golem Creations ⁶, ARRI, Fuse ⁷ and Lux Machina ⁸ together with Industrial Light and Magic's StageCraft – a virtual production filmmaking platform, and the capabilities of the Unreal Engine 4 (UE4) platform by Epic Games ⁹.

For virtual production sets to work, some numerous challenges and concerns changed the order and requirements of typical visual effects considerations. There had to be a set walkthrough in the beginning for every location before the entire set was ready for real-time performance using large environmental renders that were temporary. Virtual production sets had to be entirely created much earlier during the pre-production stage to be used digitally as a result. It was during these test runs that the production design teams brought practical set pieces that needed to be matched as well. The digital environments were also changed following the requirements of extensions and lighting adjustments that provided the final imagery. It was essential that the pre-production set-up period allowed as much room to mix and match for the practical art department to modify or change props when necessary, as well. Then

⁵ <https://www.profilestudios.com>

⁶ Golem Creations Ltd. LLC is a television production company created by Jon Favreau, an American actor and filmmaker, on August 30, 2018. "California Secretary of State". businesssearch.sos.ca.gov. Retrieved December 11, 2019.

⁷ <https://fuse-tg.com>

⁸ <https://www.luxmc.com>

⁹ <https://www.epicgames.com>

there were also technical constraints with adjusting colours that seemed relatively complex. As objects on set were lit by the digital wall environment, matching colours posed a certain degree of difficulty with some constraints. For example, if an object contained a practical on the set component as well as a digital extension on the wall, any adjustments to the light or colour on the digital wall consequentially glowed and affected the actual colours of the practical component surface. As a result, it was more convenient to modify the actual practical component instead of the virtual digital one to prevent further light contamination and unwanted errors that affected actors on the set. Volumetric 3D colour correction was done inside a spherical volume control, where any digital asset on set could be adjusted. This allowed full control of selective parts of the 3D environments that were rendered real-time by the Unreal Engine and resulted in aspects of softening or feathering for effects and colours on regions of the digital wall. The selection method for correction was done selectively using an iPad. The user chose objects or parts of them and moved the colour correction values over the volume space that were graded. Balancing the scales of certain props and objects on set in the scene was thus an issue at times. Sometimes certain on-stage props would continue into the virtual environments digitally. Other than colour correcting such digital objects, there needed to be considerations in case such objects ended up being potential unwanted light sources themselves that caused unforeseen bounce lighting on the real physical set. For example, *The Mandalorian* visual effects supervisor Richard Bluff explained that there was practical sand throughout the entire practical floor in the Mos Isley hanger scene, which then continued in the virtual 'set,' giving the sense of a much larger hanger. The practical set had yellow and red canisters for set dressing, and the scene also had

similar canisters on the digital wall that were supposed to be 9 metres beyond where the practical sand ended if seen from the camera's perspective, so those red and yellow canisters appeared smaller on the screen. Those red and yellow canisters were roughly 30 centimetres off the sand on the screens of the real stage when not looked through the lens of the camera because the screens were upright against the sandy floor. Because the LED screens provided all of the illuminations for the practical set, the problem resulted in a red and yellow light source that lit the sand in front of it. The canisters were many yards away from the sand in the camera's reality. The advantage of the LED set was that it served as the show's lighting source rather than C-stands ¹⁰ with Kinoflos ¹¹ and reflectors ¹², but it also offered an issue that contained an unintentional light source, which was an important consideration. (Seymour, M., 2020) The production process would mean the need to carefully access and balance prop management to prevent potential mistakes, especially from reflective surfaces. As for *The Mandalorian*, the Mos Isley cantina was another case of a complex mixture of prop management. The team “had to fully build out everything that appeared, in the cantina itself, which included the bar, the trim, the drinks, the dispensers, as well as everything in the middle of the bar, because they were all very reflective surfaces. Additionally, all the tables, the chairs, the booths and everything within the environment was fully built out” (Seymour, M., 2020), while the team from Industrial Light and Magic in San Francisco started to selectively turn certain

¹⁰ C-stands or “Century stands” are lighting equipment primarily used to position light modifiers, such as silks, nets, or flags, in front of light sources. (Brown, B., 2002)

¹¹ Kinoflos is the abbreviation of LED light equipment developed by manufacturer Kino Flo.

¹² Reflector(s) is a tool that helps a photographer manipulate the light by providing another surface for the light to bounce off. Definition sourced from <https://www.shawacademy.com/blog/a-beginners-guide-to-lighting-a-photo-with-a-reflector/>.

elements on the practical set to the digital set that need not be visible in the scene, but still required to cast indirect light and shadows. Essential high dynamic range (HDR) resources that are usually done during the production stage is a cornerstone of any on set visual effects team in today's workflow. However, there was no plausible sense in the Industrial Light and Magic lighting department to resort to HDRs taken from the set, since ironically the set was determined by the team's assets instead. Industrial Light and Magic had to solely rely on referenced HDRs as a result. Exterior lighting and shadows using the LED set-up did not allow for harsher shadows from a point source illumination. While the LED set provided an excellent skydome that was interactively controlled and adjusted, it was unable to produce strong lighting that created harsh shadows under environmental conditions such as direct sunlight. Therefore, traditional spotlights were still used, and the stage removed individual LED panels when required to place these spotlights. To reduce moiré patterns as an electronic artefact the LED background had to be filmed slightly out of focus, but this method made certain scenes difficult to shoot since no technique allowed defocussing requirements. There was some degree of challenge in balancing what digital objects and backgrounds were in or out of focus based on the environmental depth required in a shot. Under such circumstances, digital computer techniques were still resorted during post-production. In planning for a future enhancement for the virtual production workflow, it may be useful in some instances to have a feature found in the Unreal Engine that can detect the camera-centric distances and parameters including focal perceptions. So, the graphics engine will have a step upgrade in not only allowing correct parallax movements but also include accurate depth of field controls. There was a significant latency issue of about an 8-frame delay between

the camera information data translated into the graphical engine. The delay was further affected if the signals were required to be routed back for Steadicam functions. To combat this issue, only regions of the wall that were directly in front of the camera were rendered at a significantly higher resolution where else the areas outside of this “zone” were of lesser details. “This additional error margin gave a camera operator room to pan or tilt and not see the edge of the patch before the system could catch up with the correct field of view.” (Seymour, M., 2020) Industrial Light and Magic are beginning its service of sharing such virtual production solution across the industry and agencies under the name of StageCraft. This is a start of a new foreseeable future intended for the visual effects industry. In examining these few examples and case studies on virtual production stage workflows and processes, it is easier to assume that the visual effects pipeline will hence be very different from the typical workflow as covered and proposed earlier in this thesis. The need to facilitate more discussion and discovering the effect such new methodologies would impose into the visual effects industry is therefore perhaps a greater work in progress required during the next decade. To further add into the dilemma of the existing visual effects pipeline as a result of the available technology covered, it would seem that there is now either a merger of the post-production workflow to be part of the pre-production to allow the shoot to happen, or at the very least a requirement for certain computer-generated imagery processes to run parallel alongside pre-production to facilitate the virtual LED renders. AI could very well also be part of this process in the early stages of film production instead of just happening within post-production, resulting in another set of estimated consideration that could happen. AI systems might even pose as the next advancement in virtual processes

with the increased collaboration of existing virtual production with potential AI assistance and automation. We could perhaps see an entirely new style of film shooting process during the production stage in the future, with AI assisting as enhancements for real-time changes within the virtual production environment or on the camera that is used to film itself. Possible future AI systems may be able to help create real-time environments with the simulation of crowds and moving objects rather than a fixed stationary environment. The prospect of using highly advanced and still advancing render engines that are developed from an entirely different sector – the games industry, therefore also seem to be more conducive and accessible to the film production processes. As such, perhaps a rise of new game companies will enter the visual effects market proposing their own uniquely programmed tools and engines that serve as alternatives to the Unreal Engine by Epic Games.

Chapter 6. Conclusion

Most if not all the AI techniques mentioned thus far in this research have “not been necessarily experiencing the same level of adoption, acceptance and success... since they have not yet provided the necessary level of accuracy and ease of control.” (Sylwan, S.,2011) Every one of the visual effects pipeline steps can only begin once the previous step has been completed. Conversely, there is reasonable concern regarding the reliability that most of these algorithms produce. They are often not usable in production as possible errors might create more unwanted troublesome issues as compared to traditional means of achieving similar results. Despite the ability to generate sufficient details, the lack of control would prevent any meaningful integration into current visual effects pipeline stages. This would mean that it would still be preferred if an artist can manually produce a computer-generated mesh that is simplified but easily manipulated upon further requirements. A common issue with the research of general autonomous AI system-driven solutions is that they tend to direct amateur users who have little no prior experience in currently available tools. As such, simple solutions are generally preferred over complex ones. The more complex a solution is presented, the more likely it will affect the different departments of the visual effect pipeline and thus face a difficult acceptance into the implementation. This is “typical of the environment of the visual effects, where tools that solve some problems well and can be selected and inserted in the pipeline as needed will be preferred to more complex tools that try to solve a more generic problem formulation but provide overall less reliable results.”

(Sylwan, S.,2011) The visual effects pipeline as discussed has been commonly practised for around two decades, and as such, the digital processes and subsequent departments that have grown tremendously since then are the results of addressing solutions for the increased requirements in evolving effects in films. Often enough artists also consider if there are enough variety of differentiable parameters and modifications a visual effects tool or software can create that will be uniquely distinguishable since the algorithms would be the same across the industry. To favourably adopt certain tools, they must be a distinct advance upgrade that is a continuation for users to refine desired results instead of an autonomous solution that prevents further adjustments. In this case, it is more common for smaller commercial media companies to readily accept and adapt or try new software and tools rather than the regular big visual effect companies that have already invested heavily into existing hardware and developed working but complex production pipelines. "Visual effects companies often rely on general tools, such as Foundry's Nuke or Autodesk's Maya, for the core of their production pipeline and generally develop new solutions as plugins or add-ons to those." (Sylwan, S.,2011) Plugins allow various visual effects artist to still retain their familiarity with the proprietary software framework and user interface, and the tools used can be used above their usual production styles.

Various small teams in certain bigger visual effects company like Digital Domain and Industrial Light & Magic are often attached to the pre-production research departments to try new software that has potential uses in their working pipeline. As it is important to never hinder the actual working production environment, most of these tests are conducted in certain off-peak periods and

never during actual film production. These tools or software must usually have a common trait that is desired for the ease of production's sake – they must never destroy prior developed tools that have already proven to work for production, but instead must enable a certain form of upgrade for the current tools available, or at the very least provide a complimentary supplement for them to address complex or tedious issues that can be lessened or resolved entirely. The clear case for this would be Rotobot, as it is being marketed as a plugin. Therefore, it is of reasonable interest for independent software companies to invest efforts into developing more robust and helpful plugin solutions that might attract visual effects companies to purchase and adopt into their working pipeline and existing tools, especially if it means the lesser time needed to be set in place for their research and development. A typical research and development period would take about a year once a particular production house secures a job from a client, and this often means that the department will specifically develop tools that address unique challenges present in the visual effects shots in the project that will either serve as a unique technical achievement in the industry or provide massive ease in the workflow to have the capabilities to produce the required shots as efficiently as possible. It is not uncommon in this technological dependent industry that programmers are the main driving force needed in these various independent in-house research other than the ones we have also covered in the academic research from around the world with regards to the topic of discussion that is AI. Programmers have been around to develop the main proprietary solutions that have been adopted in the visual effects industry since the digital age and are thus still very likely to persist in their job relevancy and importance in developing the software that will include AI assistance. In most

visual effects companies, they also have specialists known as scripters that often work under a pipeline team that is meant to research and develop smaller-scale upgrades, often producing certain working macros or custom tools using the available recognised computer languages of existing production software meant to provide ease and refined control for the artists.

In retrospect, the visual effects pipeline has existed primarily under the influence of functions as dictated by production requirements. It is in this estimation of the capabilities and feasibility in the inclusion of AI that might prove rather difficult, or even in all likely hood, impossible to be part of the working pipeline simply because the pipeline has departments that serve typical filmic effect requirements such as computer-generated imagery integration onto live-action footage, creation of entire computer-generated imagery environments and assets or the need to swap a particular person or a group of actors to another character. If AI technologies are robust enough to handle all aspects of human iteration of visual effects requirements, then the pipeline itself will cease to exist. Instead, a new form of workflow would be set in place consisting of various departments related to computer science and software engineering, and perhaps only a singular creative department for design aesthetics. The result would devastate the current visual effects industry since the majority of the visual effects artists will potentially lose their career roles to industry reliant programmers. In viewing the visual effects industry as a collaborative effort with regards to commissioned outsourced productions, adding to this, a foreseeable example because of reliable AI systems would be the AI rotoscoping solution. Although it will create a cheaper option for visual effect companies to have fewer rotoscope

artists, that therein itself would cause a large proportion in the lack of rotoscope artists in the industry. A cheaper labour solution would also mean various visual effects companies that specialise in technical labour such as rotoscoping will eventually lose out or go out of business since there will be a constant offer of competitive prices lowering between them to secure a project. The visual effects departments will potentially change considering this, where rotoscoping that would ordinarily consist of around ten rotoscope artists primarily focusing just on rotoscope, there will instead be only one highly skilled specialist of computer science for AI. After covering certain departments such as the character and environmental centric AI technologies in this thesis, it is of reasonable estimation that the other various post-production processes are also involuntarily slowly being replaced in the same manner.

Overall, the pipelines will change slowly or shift into a new paradigm that will be unlike any typical and traditional workflow that visual effects depend on. A renewal usually takes place in about 10 years but will take place as an evolution of tools rather than a revolution. In addressing the concerns about human artists being replaced by AI, we should “not expect machine learning tools to replace artists soon, but rather artists can use AI tools to reduce routine tasks to focus on more creative tasks as a current situation in the industry.” (“Interview: Dmytro Korolov on the results of SIGGRAPH’s AI tools for VFX panel”, 2018) Unfortunately, “there is no denying that jobs will be displaced” (Seymour, 2018) during the next few years should AI technologies continue to develop and be successfully integrated into visual effects, similar to the reduction of special effects creators when computer-generated imagery capabilities were introduced

back during digitisation. But there are also a host of new computer solutions and programming jobs as almost all of today's visual effects are dependent on the use of computers and the available proprietary software. In understanding the impact AI systems will have in terms of career growth and changes in the industry, there will be fewer rotoscope, computer-generated character-based and environment artists in the future, and a rise of AI assistive and manipulation roles that will be highly sought after in the visual effects industry which are closely associated with fields relating to computer science and programming. Even the typical visual effects artist would have to train basic capabilities and knowledge of integrating AI into their respective work. There is a possibility that the roles will ultimately be different in certain departments, as perhaps the titles carried by these futuristic visual effects artists as well.

In the general context of this thesis where I discuss the recent developments of AI system models, resulting in my perhaps overestimations of attempting to directly apply them onto the visible issues and lack thereof present in the visual effects industry, there is something that I also believe the future will bring about great change to the way media and "effects" in general are created as commonplace gimmicks and lifestyle choices that is not only limited to the film industry as a whole. These "tools" will be able to achieve something that is commonly discussed, which is to "put visual effects in the hands of non-experts (of visual effects)". (Seymour, 2018) Mobile applications that deal with media generation and manipulation of photos and videos will easily achieve professional levels of computer-generated imagery like pixel-perfect edits, down to pixel perfect blends of fictional and live shot elements automatically via AI. As an

advancement to the technologies introduced in Chapter 4, the future appearances of believable visual effects might be readily available on personal media through smartphones as well as numerous video and photography editing applications through intuitive and simple filters and selection. Retail stores might be able to finally incorporate perfect automated full-body 3D personalised mapped system models to perfectly match and blend products such as clothes, accessories and shoes onto an Augmented Reality (AR) display with the pixel rendered and simulated “reflection” of the person in front of it, ultimately replacing normal mirrors. In fact, as akin to the case of the Zara AR app, there might be entirely new ways to incorporate these AI automated features that efficiently work on real-time data inputs to give pixel-perfect renditions for various media, that most consumer market strategies would also possibly transition and adopt them. Consumers now have a “growing appeal of synthetic characters and virtual influencers integrated into real-world settings and familiar cultural contexts, and the willingness to accept virtual personalities and even pay for entertainment by synthetic performers.” (Tercek, R., 2019) A prime example is a Japanese animated virtual singer character named Hatsune Miku from Japan, who is relatively famous as most of her major concerts are often sold out. Sapporo-based Crypton Future Media ¹ and Yamaha created a vocal synthesizer voicebank which they dubbed as a Vocaloid for marketing purposes, where Miku first appeared. Her name means “first sound of the future”. The stage setup typically consists of the use of multiple projectors, and a couple of sheets of polyethene to give the holographic impression which covers the front and back simultaneously. The

¹ <https://www.crypton.co.jp>

prospects of generating more varied distinct and realistic looking virtual characters are viable, as visual effects no longer just retain for a film on the screen on a wall, but in a real-world setting like a sort of holographic projection. Current game engines as well as ongoing research of generative adversarial networks (GANs) that are capable of generating realistic motion data, characters and environments without any prior input are in a very early stage of development but will be able to assist smaller media creation teams or individuals greatly in creating personal filmic media content where manpower or tedious tasks would foster challenges. Disney Research and Rutgers ² University published a project which the AI software is used “to generate rapid prototypes: the Disney AI can interpret natural language descriptions of scenes and settings in movie screenplays to generate storyboards and rough animation sequences.” Specifically, “their text-to-animation model produces animations without the need for annotated data or a pretraining step but relying on given input text describing certain activities. Then, a pipeline dubbed Cardinal takes as input the actions and creates previsualizations in Unreal Engine. Drawing on the predefined animations library, pre-uploaded objects, and models which it can use to create characters, it produces 3D animated videos that approximately depict the processed script.” (Wiggers, K., 2019) The capabilities of AI systems might also further reduce the line between entertainment platforms of games and films, as visual and graphical capabilities start to diminish into the realm of hyper-realism. Just as game engines are used in the latest AI systems for virtual production, instead, AI visual effects solutions and film direction can be crossed into game production. At the

² <https://www.rutgers.edu>

cornerstone beginning of this field, California-based start-up rct studio ³ “is leveraging cutting-edge AI in its quest for a movie experience that is both interactive and immersive.” (Sarazen, M., 2019) The aim of rct studio is to create an environment in virtual reality that contains non-player characters (NPCs) that can be interacted with and yet behave uniquely in the same manner as other human users. The company created the Morpheus engine that can learn and translate screenplay texts consisting of objects, dialogue, action physics and character movements automatically into a full 3D real-time simulation. Founder of rct studio, Jesse Lyu explains that they wish to create user interactions that will allow infinite endings to carry out. Current AI systems such as machine learning can let NPCs learn and train through numerous simulations and thus consequentially be adapted to various unspecified situations. He believes that AI can learn well from films as they often come with well thought out definitions while exploring numerous character-centric emotional factors and actions. “Take the most complex movies – such as the Harry Potter or the Marvel series — as examples. While these movies have so many characters, they still have a clear storyline, a clear in-and-out. There are a lot of standardized, industry-grade screenplay scripts that data-driven machines can utilize and learn from.” As a result of this project, the Morpheus engine has caught the attention of several media agencies and studios. Jesse Lyu “estimates the engine could save a typical Hollywood action film up to US\$20 million by automating key processes”, proving its cost efficiency. Canada-based creative director Sean MacPhedran, who is also an Ottawa International Gamer Conference co-founder, mentions that “it would

³ <https://rct.ai/en-us>

definitely be a time-saver in creating storyboards, understanding the tempo of a story – in effect it sounds like the tool creates a very rough cut of the narrative.” (Sarazen, M., 2019) When we go back to films or more specifically the visual effects industry – the resulting trends of production might never change. The limitations to the quality of small production visual effects might still be easily distinguishable from high budget productions in major visual effects companies where more specially AI or visual effects trained departments work on big films. Despite this, smaller studios and individual users will be able to effectively achieve the desired quality of visual effects in various media when labour and technical challenges are present due to the assistance of the discussed AI tools. It will certainly take an amalgamation of factors during the development of various AI algorithms and training data to reach a point of base-level “perfection” in naturally automated realism of virtual effects. Even that, I would assume is not an impossible eventuality, resulting in only subtle differences in high-quality filmic media from all forms of artists and the big visual effects companies. There will be of reasonable concern then if all film media of the future are fully created by AI. What will be left for visual effects artists? Lev Manovich, a leading theorist who addresses many new media theories, suggests that “Many modern thinkers and artists have envisioned a future where humans, liberated by machines from mechanical and boring work, will be engaging only in play and art. But if automation of production by AI continues, eventually it will be these AI playing and making art.” (Manovich, L., 2018) Assuming that AI will advance in development for visual effects, it is therefore important for the visual effects industry to find a balance in using AI with visual effects artists’ authentic skillsets and knowledge of technologies. Like all forms of art, the quality of the film

industry comes down to a function of skill and time. By exploring how AI can overall effectively shorten the required time for visual effects production in some ways, I hope that the technology will bring about significant solutions to the current issues for the artists in the industry.

During this point of discussion, I believe that there will be a variation of AI systems that will directly enhance or replace current visual effect pipeline processes in some manner or another. This investigation and record of emerging new technology and its attempt to bridge towards visual effects are by no means accomplished, as the entire field of study and practical applications are only in the beginnings. I hope that the research has allowed a basic coverage of AI systems, visual effects methodologies and a general understanding of difficulties faced in the current industry production workflows as part of the growing history of film. As a result, it could very well be that the visual effects pipeline we are familiar with today, will soon no longer exist. Instead, a new evolutionary system workflow that would be shared from across all forms of media entertainment would be developed, under the assistance of AI, with the human touch.

Appendix A: Interview Transcripts

The interviews were conducted by me as the primary researcher on this paper using an online google survey format. There was a total of 6 individuals with various visual effects backgrounds and experiences that responded to the survey. The survey was conducted to gather a few perspectives on A.I in the visual effects industry, whether artists are informed and are aware of the emergence and possible integration to the current workflows. Below are the full transcriptions of the interviews.

INTERVIEW 1:

Benjamin Seide, Associate Professor, Nanyang Technological University Submitted on 18/12/2019.

1. Are you (or your team) already using A.I. tools in your VFX production?

Answered: No

2. What are some of the A.I tools you (or your team) have used, and what are your thoughts about them?

Answered:

3. If you haven't used any A.I tools, which ones have you heard of, and what are your thoughts?

Answered: *Rotobot* but haven't used

4. Rotoscoping is something everyone expects A.I to help in the VFX industry.

Have you (or your team) tried A.I rotoscoping, and if so, what are your impressions?

Answered: Haven't tried

5. When do you think "Autonomous Rotoscoping" will be effective as a fully functioning feature on the production standard?

Answered: 2022

6. "Deep Fakes" are the new A.I facial replacement technology we have present in some mobile apps, instances of videos on pop culture icons as well as adult videos. What are your thoughts feelings about the technology, and do you expect it to be available for Digital VFX?

Answered: Scary, can't trust any video anymore. Expect Visual Effects tools in 2022

7. With the recent rising awareness and interest with A.I research and development for Digital VFX, how do you think A.I will impact VFX industry of the future?

Answered: Reduce labour hours, more hobbier Visual Effects

8. In the future, what do you think is the general expectation of the standard of VFX with the successful inclusion of A.I?

Answered: Pessimistically Visual Effects will be seen even more as a commodity.

Optimistically, indie filmmakers will have possibilities held previously only by big budget productions

INTERVIEW 2:

Professor, Submitted on 11/02/2020.

1. Are you (or your team) already using A.I. tools in your VFX production?

Answered: No

2. What are some of the A.I tools you (or your team) have used, and what are your thoughts about them?

Answered:

3. If you haven't used any A.I tools, which ones have you heard of, and what are your thoughts?

Answered: *Open Pose, deepart.io*

4. Rotoscoping is something everyone expects A.I to help in the VFX industry. Have you (or your team) tried A.I rotoscoping, and if so, what are your impressions?

Answered: yes. we tested *open pose*. It is partly applicable, but not without manual finetuning. Not usable for movie productions at the moment.

5. When do you think “Autonomous Rotoscoping” will be effective as a fully functioning feature on the production standard?

Answered: It will take some more years certainly. Probably soon employable for standard cases for basic operations. Special cases might remain manufacturing work.

6. “Deep Fakes” are the new A.I facial replacement technology we have present in some mobile apps, instances of videos on pop culture icons as well as adult videos. What are your thoughts feelings about the technology, and do you expect it to be available for Digital VFX?

Answered: For digital VFX certainly, because the apps are already operating for digital visual effects. The matter is the quality and believability. I have seen quite convincing examples already. In my opinion, Gaming, AR, VR will benefit from it first.

7. With the recent rising awareness and interest with A.I research and development for Digital VFX, how do you think A.I will impact the VFX industry of the future?

Answered: A lot.

8. In the future, what do you think is the general expectation of the standard of VFX with the successful inclusion of A.I?

Answered: More productivity with lower budgets. Less labor work. Loss of jobs. More creative chances.

INTERVIEW 3:

Christian Pundschus, Freelance VFX Supervisor/Comp Lead, Submitted on 11/02/2019.

1. Are you (or your team) already using A.I. tools in your VFX production?

Answered: Yes

2. What are some of the A.I tools you (or your team) have used, and what are your thoughts about them?

Answered: Style Transfer

3. If you haven't used any A.I tools, which ones have you heard of, and what are your thoughts?

Answered:

4. Rotoscoping is something everyone expects A.I to help in the VFX industry. Have you (or your team) tried A.I rotoscoping, and if so, what are your impressions?

Answered: Tried it. Good for rough rotos and garbage mattes. Still not quite good enough to replace established manual techniques (e.g. Silhouette)

5. When do you think "Autonomous Rotoscoping" will be effective as a fully functioning feature on the production standard?

Answered: 5 Years

6. "Deep Fakes" are the new A.I facial replacement technology we have present in some mobile apps, instances of videos on pop culture icons as well as adult videos. What are your thoughts feelings about the technology, and do you expect it to be available for Digital VFX?

Answered: As of now, classical deep fakes hold up pretty well on highly compressed video (Youtube) where pores and small wrinkles are not visible anyway. A.I. already plays a major role in enhancing CG faces (see "Smart Hulk") and the borders to deepfakes and similar applications will eventually disappear.

7. With the recent rising awareness and interest with A.I research and development for Digital VFX, how do you think A.I will impact the VFX industry of the future?

Answered: Massively. In a few years, AI will assist VFX in every aspect.

8. In the future, what do you think is the general expectation of the standard of VFX with the successful inclusion of A.I?

Answered: More realism for a lower price.

INTERVIEW 4:

Eran Dinur, VFX Supervisor, Author, Teacher, Submitted on 11/02/2019.

1. Are you (or your team) already using A.I. tools in your VFX production?

Answered: No

2. What are some of the A.I tools you (or your team) have used, and what are your thoughts about them?

Answered:

3. If you haven't used any A.I tools, which ones have you heard of, and what are your thoughts?

Answered: Particularly interested in AI roto tools, which are still not accurate enough, and AI reconstruction tools, which are getting seriously impressive.

4. Rotoscoping is something everyone expects A.I to help in the VFX industry. Have you (or your team) tried A.I rotoscoping, and if so, what are your impressions?

Answered:

5. When do you think "Autonomous Rotoscoping" will be effective as a fully functioning feature on the production standard?

Answered: Still not accurate enough for VFX, but looking promising!

6. "Deep Fakes" are the new A.I facial replacement technology we have present in some mobile apps, instances of videos on pop culture icons as well as adult videos. What are your thoughts feelings about the technology, and do you expect it to be available for Digital VFX?

Answered: This is an area that is making very impressive strides forward. I believe it will become immensely useful in the near future.

7. With the recent rising awareness and interest with A.I research and development for Digital VFX, how do you think A.I will impact the VFX industry of the future?

Answered: I am hoping that within a few years some of these new tools will be accurate and reliable enough to be used in VFX regularly. That said, VFX is still largely a problem-solving process that needs human artistry and creativity.

8. In the future, what do you think is the general expectation of the standard of VFX with the successful inclusion of A.I?

Answered: See previous answer.

INTERVIEW 5:

Daniel Stern, FX Lead, The Mill LA, Submitted on 13/02/2019.

1. Are you (or your team) already using A.I. tools in your VFX production?

Answered: No

2. What are some of the A.I tools you (or your team) have used, and what are your thoughts about them?

Answered:

3. If you haven't used any A.I tools, which ones have you heard of, and what are your thoughts?

Answered: Face replacement, volume simulation, auto roto, auto depth, auto forest, sounds all very interesting. issues occur if the AI tool delivers a decent but not perfect (=final) quality and the results have to be enhanced after. E.g.: Face replacement of an actor, the quality is washy and not sharp enough. Not much to do about it: not usable.

4. Rotoscoping is something everyone expects A.I to help in the VFX industry. Have you (or your team) tried A.I rotoscoping, and if so, what are your impressions?

Answered: Not tried

5. When do you think “Autonomous Rotoscoping” will be effective as a fully functioning feature on the production standard?

Answered: 1-2 years

6. “Deep Fakes” are the new A.I facial replacement technology we have present in some mobile apps, instances of videos on pop culture icons as well as adult videos. What are your thoughts feelings about the technology, and do you expect it to be available for Digital VFX?

Answered: Yes I do. But as mentioned before these solutions have the problem that they can't be polished easily afterwards.

7. With the recent rising awareness and interest with A.I research and development for Digital VFX, how do you think A.I will impact the VFX industry of the future?

Answered: It won't replace it. It'll help though.

8. In the future, what do you think is the general expectation of the standard of VFX with the successful inclusion of A.I?

Answered: I think that the current standard of high-end VFX (features / commercials) can already reach a photoreal quality.

I don't think that AI will help to make things more photoreal. But it'll help to get there easier.

INTERVIEW 6:

Christian Kalata, Head of CG, MPC, Submitted on 04/03/2019.

1. Are you (or your team) already using A.I. tools in your VFXproduction?

Answered: No

2. What are some of the A.I tools you (or your team) have used, and what are your thoughts about them?

Answered:

3. If you haven't used any A.I tools, which ones have you heard of, and what are your thoughts?

Answered: nvidias gaussian

4. Rotoscoping is something everyone expects A.I to help in the VFX industry.

Have you (or your team) tried A.I rotoscoping, and if so, what are your impressions?

Answered: tried *mocha*. quite okay but not accurate enough for the footage we have to deal with

5. When do you think “Autonomous Rotoscoping” will be effective as a fully functioning feature on the production standard?

Answered: not in the immediate future

6. “Deep Fakes” are the new A.I facial replacement technology we have present in some mobile apps, instances of videos on pop culture icons as well as adult videos. What are your thoughts feelings about the technology, and do you expect it to be available for Digital VFX?

Answered: exciting. takes some years to get there until perfection as it is tied to a lot of restrictions in the footage

7. With the recent rising awareness and interest with A.I research and development for Digital VFX, how do you think A.I will impact the VFX industry of the future?

Answered: It won't revolutionize things in VFX for quite some time. I think it will tie in with real time solutions. VFX will remain pretty much the same in the next 5 years.

8. In the future, what do you think is the general expectation of the standard of VFX with the successful inclusion of A.I?

Answered: flawless results are paramount. anything less will result in the contemporary way of approaching problems

Bibliography:

Adee, S. (2020, April 26). *What Are Deepfakes and How Are They Created?*

Retrieved May 06, 2020, from <https://spectrum.ieee.org/tech-talk/computing/software/what-are-deepfakes-how-are-they-created>

Anderson, M. (n.d.). *A comprehensive guide to the state-of-art in how AI is transforming the visual effects (VFX) industry*. Retrieved October 16, 2019, from <https://rossdawson.com/futurist/implications-of-ai/comprehensive-guide-ai-artificial-intelligence-visual-effects-vfx/>

Anokhin, I., Solovev, P., Korzhenkov, D., Kharlamov, A., Khakhulin, T., Silvestrov, A., ... & Sterkin, G. (2020). *High-Resolution Daytime Translation Without Domain Labels*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7488-7497).

Alexander, J. (2015, October 20). *Furious 7 used 350 CGI shots of Paul Walker*.

Retrieved January 06, 2020, from <https://www.polygon.com/2015/10/20/9577863/furious-7-used-350-cgi-shots-of-paul-walker>

AI-powered 3D animation. (n.d.). Retrieved December 03, 2020, from <https://getrad.co/>

AI (Artificial Intelligence). (n.d.). Retrieved August 06, 2020, from <https://exploreai.org/p/ai-definition>

Badr, W. (2019, July 01). *Auto-Encoder: What Is It? And What Is It Used For?* (Part 1). Retrieved August 15, 2019, from <https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for-part-1-3e5c6f017726>

Bau, D., Strobel, H., Peebles, W., Zhou, B., Zhu, J. Y., & Torralba, A. (2020). *Semantic photo manipulation with a generative image prior*. arXiv preprint arXiv:2005.07727.

Brownlee, J. (2019, July 19). *A Gentle Introduction to Generative Adversarial Networks (GANs)*. Retrieved August 15, 2019, from <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>

Brown, Blain (2002). *Cinematography: imagemaking for cinematographers, directors and videographers*. Focal Press. p. 252. ISBN 0-240-80500-3.

Chakraborty, A. (2019, June 05). *What is Semantic Segmentation?* Retrieved February 05, 2020, from <https://www.learnopencv.com/pytorch-for-beginners-semantic-segmentation-using-torchvision/>

Choi, Y., Uh, Y., Yoo, J., & Ha, J. W. (2020). *Stargan v2: Diverse image synthesis for multiple domains*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8188-8197).

Chu, M., Xie, Y., Mayer, J., Leal-Taixé, L., & Thurey, N. (2020). Learning temporal coherence via self-supervision for GAN-based video generation. *ACM Transactions on Graphics (TOG)*, 39(4), 75-1.

Cram, C. (2012). *Digital Cinema: The Role of the Visual Effects Supervisor*. Film History, (Vol. 24).

Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I. (2016).

Curtin, M., & Vanderhoef, J. (2015). A Vanishing Piece of the Pi: The Globalization of Visual Effects Labor. *Television & New Media*, 16(3), 219–239.
<https://doi.org/10.1177/1527476414524285>

Daly, C. (2017, June 11). Augmented Reality Needs AI In Order To Be Effective. Retrieved March 05, 2020, from
https://aibusiness.com/document.asp?doc_id=760453

Edwards, G. (2015, July 1). *VFX ABC: M is for Matte Painting*. Retrieved August 23, 2020, from <https://www.tested.com/art/movies/532866-m-matte-painting/>

Failes, I. (2019, May 11). The New Artificial Intelligence Frontier of VFX. Retrieved April 04, 2020, from <https://www.vfxvoice.com/the-new-artificial-intelligence-frontier-of-vfx/>

Feinstein, L. (2019, July 07). *Deepfakes: The art of deception*. Retrieved February 08, 2020, from <https://a-d-o.com/journal/deepfakes>

Garbage Matte. (2019, February 06). Retrieved August 06, 2019, from <https://moviola.com/visual-glossary/garbage-matte/>

Genova, K., Cole, F., Maschinot, A., Sarna, A., Vlastic, D., & Freeman, W.T. (2018).

Unsupervised Training for 3D Morphable Model Regression. 2018 IEEE/CVF

Conference on Computer Vision and Pattern Recognition, 8377-8386.

Giardina, C. (2014, February 27). Revealing 'Rhythm & Hues: Life After Pi'

Doc Exposes Grief, Anger and Troubled Business (Video). Retrieved November 06,

2019, from <https://www.hollywoodreporter.com/behind-screen/revealing-rhythm-hues-life-pi-682526>

Giardina, C. (2015, December 11). How 'Furious 7' Brought the Late Paul Walker

Back to Life. Retrieved June 14, 2019, from

<https://www.hollywoodreporter.com/behind-screen/how-furious-7-brought-late-845763>

Gituma, M. (2019, June 23). What is Optical Flow and why does it matter in deep

learning. Retrieved July 07, 2019, from [https://medium.com/swlh/what-is-](https://medium.com/swlh/what-is-optical-flow-and-why-does-it-matter-in-deep-learning-b3278bb205b5)

[optical-flow-and-why-does-it-matter-in-deep-learning-b3278bb205b5](https://medium.com/swlh/what-is-optical-flow-and-why-does-it-matter-in-deep-learning-b3278bb205b5)

Harris, S. A. K. (2019, June 21). *Technology of the future: Could deepfakes replace*

VFX? Retrieved October 17, 2019, from [https://uk.movies.yahoo.com/technology-](https://uk.movies.yahoo.com/technology-of-the-future-will-deepfakes-replace-visual-effects-120710805.html)

[of-the-future-will-deepfakes-replace-visual effects-120710805.html](https://uk.movies.yahoo.com/technology-of-the-future-will-deepfakes-replace-visual-effects-120710805.html)

Hodge, S. (2018). *Automation of Rough Rotoscopy*. Accessed on 2 August 2019,

from [https://kognat.com/wp-content/uploads/2018/09/roto-bot-ai-based-](https://kognat.com/wp-content/uploads/2018/09/roto-bot-ai-based-masks-from-kognat-software.pdf)

[masks-from-kognat-software.pdf](https://kognat.com/wp-content/uploads/2018/09/roto-bot-ai-based-masks-from-kognat-software.pdf)

Itzkoff, D. (2016, December 27). How 'Rogue One' Brought Back Familiar Faces. The New York Times. <https://www.nytimes.com/2016/12/27/movies/how-rogue-one-brought-back-grand-moff-tarkin.html>.

Insta360 ONE R Receives Major Update With AI-Powered Features. (2020, September 18). Retrieved November 15, 2020, from <https://blog.insta360.com/one-r-massive-update/>

Interview: Dmytro Korolov on the results of SIGGRAPH's AI tools for VFX panel. (2018). Retrieved 30 July 2019, from <https://3deeplearner.com/dmytro-siggraph-ai-tools-for-visual-effects-panel/>

Introducing the NVIDIA RTX Ray Tracing Platform. (2019, July 29). Retrieved August 15, 2019, from <https://developer.nvidia.com/rtx/raytracing>

Jesus, A. (2019, November 22). Virtual Mirrors and Computer Vision - 9 Current Applications. Retrieved August 15, 2020, from <https://emerj.com/ai-sector-overviews/virtual-mirrors-and-computer-vision/>

Karacan, L., Akata, Z., Erdem, A., & Erdem, E. (2018). *Manipulating attributes of natural scenes via hallucination*. arXiv preprint arXiv:1808.07413.

Kaufman, D. (2019). *Global VFX: State of the Industry 2019*. Retrieved 30 July 2019, from <http://visual-effectsvoice.com/global-visual-effects-state-of-the-industry-2019/>

Keil, C. and Whissel, K. (2016). *Editing and Special/Visual Effects: Behind the Silver Screen: A Modern History of Filmmaking*. New Brunswick, New Jersey: Rutgers University Press.

Kim, H., Garrido, P., Tewari, A., Xu, W., Thies, J., Niessner, M., Pérez, P., Richardt, C., Zollhöfer, M., & Theobalt, C. (2018). *Deep video portraits*. ArXiv, abs/1805.11714.

Kim, P. (2017). *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*: Apress.

Kimmel, R., Klette, R., & Sugimoto, A. (2011). *Computer Vision - ACCV 2010* (Vol. 6492).

Laffont, P. Y., Ren, Z., Tao, X., Qian, C., & Hays, J. (2014). *Transient attributes for high-level understanding and editing of outdoor scenes*. *ACM Transactions on graphics (TOG)*, 33(4), 1-11.

Liu, Y. L., Lai, W. S., Yang, M. H., Chuang, Y. Y., & Huang, J. B. (2020). *Learning to See Through Obstructions*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14215-14224).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. *Nature*, 521, 436-444. doi:10.1038/nature14539

Levine, N. (2019, June 30). *A Gross App Used To 'Undress' Women Has Been Shut Down*. Retrieved March 08, 2020, from <https://www.refinery29.com/en-gb/2019/06/236659/deepnude-app-shut-down>

- Liu, M., Breuel, T., & Kautz, J. (2017). *Unsupervised Image-to-Image Translation Networks*. ArXiv, abs/1703.00848.
- Liu, D., Jiang, Y., Pei, M., & Liu, S. (2018). *Emotional image colour transfer via deep learning*. Pattern Recognition Letters, 110, 16-22.
- Lombardi, S., Simon, T., Saragih, J., Schwartz, G., Lehrmann, A., & Sheikh, Y. (2019). *Neural volumes: Learning dynamic renderable volumes from images*. arXiv preprint arXiv:1906.07751.
- Mallick, S. (2018, November 05). *Image Segmentation*. Retrieved March 05, 2020, from <https://www.learnopencv.com/image-segmentation/>
- Marr, B. (2019, October 04). The Magic Of Smart Mirrors: Artificial Intelligence, Augmented Reality And The Internet of Things. Retrieved July 12, 2020, from <https://www.forbes.com/sites/bernardmarr/2019/10/04/the-magic-of-smart-mirrors-artificial-intelligence-augmented-reality-and-the-internet-of-things/?sh=77d63939615b>
- Marsh, C. (2018, May 30). *VFX Companies Struggle Even as Their Movies Break Records*. Retrieved August 14, 2019, from <https://variety.com/2018/artisans/production/visual-effects-summer-box-office-1202824543/>
- Manovich, L. (2018). *AI Aesthetics*. Moscow: Strelka Press.

M. Mori, K. F. MacDorman and N. Kageki. (2012, June) *The Uncanny Valley [From the Field]*. IEEE Robotics & Automation Magazine, 19(2), 98-100, doi: 10.1109/MRA.2012.2192811.

Medusa. (n.d.). Retrieved January 05, 2021, from <https://studios.disneyresearch.com/medusa/>

Montgomery, J. (2018, July 18). *MPC Replicating Rachael in Blade Runner 2049* (updated). Retrieved October 20, 2019, from <https://www.fxguide.com/fxfeatured/mpc-replicating-rachael-in-blade-runner-2049/>.

Mozar, A. (2016, January 9). *The Benefits and Drawbacks of Chroma Shooting for VFX*. Retrieved January 08, 2020, from <https://thevirtualassist.net/benefits-and-drawbacks-of-chroma-shooting/>

Murphy, H. (2019, August 16). *Cyber security companies race to combat 'deepfake' technology*. Retrieved February 08, 2020, from <https://www.ft.com/content/63cd4010-bfce-11e9-b350-db00d509634e>

Neekhara, P., Hussain, S., Jere, M., Koushanfar, F., & McAuley, J. (2020). *Adversarial Deepfakes: Evaluating Vulnerability of Deepfake Detectors to Adversarial Examples*. arXiv preprint arXiv:2002.12749.

Okun, J. A., Zwerman, S., & Visual Effects Society. (2010). *The VES handbook of visual effects: Industry standard VFX practices and procedures*. Amsterdam: Focal Press.

- Patruno, A. (2017, February 15). *VFX and the Perceptual Realism of Cinema*. Retrieved August 15, 2020, from <https://medium.com/luci-scribia/vfx-and-the-perceptual-realism-of-cinema-19dda845f0cc>
- Pangburn, D. J. (2019, September 23). *You've been warned: Full body deepfakes are the next step in AI-based human mimicry*. Retrieved October 17, 2019, from <https://www.fastcompany.com/90407145/youve-been-warned-full-body-deepfakes-are-the-next-step-in-ai-based-human-mimicry>.
- Park, S., Ryu, H., Lee, S., Lee, S., & Lee, J. (2019). Learning predict-and-simulate policies from unorganized human motion data. *ACM Transactions on Graphics (TOG)*, 38(6), 1-11.
- Pennington, A. P. (2019, July 04). How AI is reinventing visual effects. Retrieved May 5, 2020, from <https://www.ibc.org/trends/how-ai-is-reinventing-visual-effects/4060.article>
- Pidhorskyi, S., Adjero, D. A., & Doretto, G. (2020). *Adversarial Latent Autoencoders*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14104-14113).
- Prince, S. (2011). *Digital Visual Effects in Cinema: The Seduction of Reality*. New Brunswick, New Jersey: Rutgers University Press.
- Read, P., In Meyer, M.-P., & Gamma Group. (2000). *Restoration of motion picture film*. Oxford: Elsevier Science.

Robertson, B. (2018, May 07). Sympathy for the Devil: How Digital Domain Helped Create Thanos for 'Avengers: Infinity War'. Retrieved March 16, 2020, from <https://www.awn.com/vfxworld/sympathy-devil-how-digital-domain-helped-create-thanos-avengers-infinity-war>

Sarazen, M. (2019, April 17). Direct and Star in Your Own Movie With California AI Startup Rct Studio. Retrieved September 23, 2020, from <https://syncedreview.com/2019/04/17/direct-and-star-in-your-own-movie-with-california-ai-startup-rct-studio/>

Sengupta, S., Jayaram, V., Curless, B., Seitz, S., & Kemelmacher-Shlizerman, I. (2020). *Background Matting: The World is Your Green Screen*, arXiv:2004.00626. Retrieved from <https://ui.adsabs.harvard.edu/abs/2020arXiv200400626S>

Seymour, M. (2011, October 10). The Art of Roto: 2011. Retrieved May 05, 2020, from <https://www.fxguide.com/fxfeatured/the-art-of-roto-2011/>

Seymour, M. (2018). 'AI.' for VFX at SIGGRAPH Part 1. Retrieved 30 July 2019, from <https://www.fxguide.com/fxfeatured/a-i-for-visual-effects-at-siggraph-part-1/>

Seymour, M. (2018, November 30). Rotobot, bringing Machine Learning to Roto. Retrieved October 25, 2019, from <https://www.fxguide.com/quicktakes/rotobot-bringing-machine-learning-to-roto/>

Seymour, M. (2020, March 04). Art of LED Wall Virtual Production, Part One: 'Lessons from the Mandalorian'. Retrieved August 5, 2020, from

<https://www.fxguide.com/fxfeatured/art-of-led-wall-virtual-production-part-one-lessons-from-the-mandalorian/>

Sharma, R. (2020, July 18). *Self-supervised Attention Mechanism for Dense Optical Flow Estimation*. Retrieved July 19, 2020, from <https://towardsdatascience.com/self-supervised-attention-mechanism-for-dense-optical-flow-estimation-b7709af48efd>

Shoard, C. (2016, December 21). Peter Cushing is dead. Rogue One's resurrection is a digital indignity | Catherine Shoard. The Guardian. <https://www.theguardian.com/commentisfree/2016/dec/21/peter-cushing-rogue-one-resurrection-cgi>.

Stojković, N. (2020, January 28). *Introduction To Green Screen, Blue Screen, And Chroma Keying*. Retrieved February 07, 2020, from <https://filmdaft.com/introduction-to-green-screen-blue-screen-and-chroma-keying/>

Sylwan, S. (2010). The application of vision algorithms to visual effects production. Paper presented at the Proceedings of the 10th Asian conference on Computer vision - Volume Part I, Queenstown, New Zealand.

Tercek, R. (2019, June 05). Artificial Intelligence is completely reinventing media and marketing. Retrieved October 08, 2020, from <https://medium.com/id-in-the-iot/artificial-intelligence-is-completely-reinventing-media-and-marketing-d724c150ece3>

The Mill. (n.d.). The Human Race - Behind The Scenes. Retrieved November 10, 2020, from <http://archive.themill.com/portfolio/3517/the-human-race---behind-the-scenes->

Thompson, Z. (2017, December 23). Digitally Reviving Peter Cushing For Rogue One Is Disrespectful. HuffPost. https://www.huffpost.com/entry/digitally-reviving-peter-cushing_b_13800074.

The Benefits and Drawbacks of Chroma Shooting for VFX. (2018, September 14). Retrieved May 15, 2019, from <https://thevirtualassist.net/benefits-and-drawbacks-of-chroma-shooting/?msg=fail>

Tron (2010). (n.d.). Retrieved September 23, 2020, from <https://computeranimationhistory-cgi.jimdofree.com/tron-2010/>

VFX in the modern age. (2017). Retrieved 30 July 2019, from https://www.foundry.com/insights/film-tv/VISUAL_EFFECTS-in-the-modern-age

Venkatasawmy, R. (2013). *The digitisation of cinematic visual effects: Hollywood's coming of age*. Lexington Books.

Venkatasawmy, R. (2016). *Crisis in the digital visual effects industry and Hollywood cinema*. Film International Vol. 14, Issue 1, p. 6-23. Intellect.

Visual effects. (n.d). In Wikipedia. Retrieved August 14, 2019, from https://en.wikipedia.org/wiki/Visual_effects

Wang, L., Chen, W., Yang, W., Bi, F., & Yu, F. R. (2020). A State-of-the-Art Review on Image Synthesis With Generative Adversarial Networks. *IEEE Access*, 8, 63514-63537. doi:10.1109/ACCESS.2020.2982224

What is Artificial Intelligence (AI)? (n.d.). Retrieved September 06, 2020, from <https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence>

What is Rotoscoping? (n.d.). Retrieved September 06, 2019, from <https://www.futurelearn.com/info/courses/vfx-for-filmmakers/0/steps/13255>

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*. *International Data Privacy Law*, 7(2), 76-99. doi:10.1093/idpl/idx005

Walker, J. (n.d.). *The artistry of matte painting*. Retrieved June 14, 2020, from <https://www.futurelearn.com/courses/vfx-for-filmmakers/0/steps/13286>

Williams, D. E. (2006). *Symbolic Victory*. American Cinematographer.

Wiggers, K. (2019, April 12). Disney's AI generates storyboard animations from scripts. Retrieved February 17, 2020, from <https://venturebeat.com/2019/04/12/disneys-ai-generates-storyboard-animations-from-screenplays/>

Whitehurst, A. (n.d.). The Visual Effects Pipeline. Retrieved December 16, 2018, from <http://www.andrew-whitehurst.net/pipeline.html>

Zara's Augmented Reality App Brings Virtual Models to Life in Stores. (n.d.).

Retrieved October 02, 2020, from <https://futurestoreseast.wbresearch.com/zara-augmented-reality-app-virtual-model-strategy-ty-u>

Zhang, X. C., Tsai, Y. T., Pandey, R., Zhang, X., Ng, R., & Jacobs, D. E. (2020). *Portrait Shadow Manipulation*. arXiv preprint arXiv:2005.08925.

Zakharov, E., Shysheya, A., Burkov, E., & Lempitsky, V. (2019). *Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*.