

Academic Article

Generative Adversarial Network สำหรับการสร้างตัวละคร DOTA2

Generative Adversarial Network for DOTA2 Character Generation

สรรัชย์ อังควินิจวงศ์^{1*} พงศรัษฎ์ ชูทัย^{2*} กฤตติพัฒน์ ชื่นพิทยาธุ^{3*} ธัชช นาคเสนีย์^{4*} จูติรัตน์ สิริบวรรัตนกุล^{5*}
Saranchai Angkawinijwong^{1*} Pongsarat Chootai^{2*} Krittipat Chuenphitthayavut^{3*} Theethut Narksenee^{4*}
Thitirat Siriborvornratanakul^{5*}

*คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ 148 หมู่3 ถนนเสรีไทย คลองจั่น บางกะปิ กรุงเทพฯ 10240

*Graduate School of Applied Statistics, National Institute of Development Administration, 148 Serithai Road,
Klong-Chan, Bangkok, Bangkok Thailand 10240

¹E-mail: saranchai.ang@stu.nida.ac.th; ²E-mail: pongsarat.cho@stu.nida.ac.th;

³E-mail: krittipat@g.swu.ac.th; ⁴E-mail: theethut.nar@stu.nida.ac.th ⁵E-mail: thitirat.s@nida.ac.th;

Received: xx/yy/20zz; Revised: xx/yy/20zz; Accepted: xx/yy/20zz

เกม DOTA2 เป็นเกมในอุตสาหกรรม E-Sport Game ที่ได้รับความนิยม และมีมูลค่าทางการตลาดสูง โดยเฉพาะถ้าพัฒนาลักษณะตัวละครในเกมให้มีความหลากหลายสามารถเพิ่มรายได้ให้กับบริษัท และความพึงพอใจของผู้เล่นได้ โดยจากหลายงานวิจัยสามารถแสดงถึงประสิทธิภาพของการประยุกต์ใช้ Generative Adversarial Networks (GAN) มาใช้ในการสร้างภาพการ์ตูน หรือสร้างภาพบุคคลขึ้นใหม่อย่างดี งานวิจัยนี้จึงมีวัตถุประสงค์ต้องการศึกษาประสิทธิภาพของการประยุกต์ใช้ Style-Gan เพื่อใช้ในการสร้างและออกแบบตัวละคร DOTA2 ขึ้นมาใหม่ ในลักษณะการสร้างภาพชื่อจำนวนน้อยผ่านการได้ตอบข้ามโดเมน (Few-shot Image Generation via Cross-domain Correspondence) ซึ่งมีจุดเด่นคือ ไม่ต้องการจำนวนรูปภาพที่ใช้เทรนจำนวนมากเหมือนการเทรน GAN ทั่วไปและไม่เกิด Overfitting ที่มีความถูกต้องสูงเกินไป โดยข้อมูลที่ใช้ในการเรียนรู้ คือ รูปขนาดย่อของตัวละครฮีโร่เกม DOTA2 ขนาด 256x144 พิกเซล จำนวน 111 รูป และ สร้างขึ้นจาก 3 โมเดล คือ 1) โมเดล Source_ffhq (โมเดล styleGAN2 ที่เรียนรู้สร้างภาพใบหน้าคน), 2) โมเดล Church (โมเดล styleGAN2 ที่เรียนรู้สร้างภาพโบสถ์) และ 3) โมเดล Horses (โมเดล styleGAN2 ที่เรียนรู้สร้างภาพม้า) โดยผลการเปรียบเทียบในเชิงปริมาณพบว่า โมเดลที่ประยุกต์มาจากโมเดล FFHQ แสดงให้เห็นว่า โมเดลสามารถดัดแปลงส่วนตาและจมูกของคน มาเป็นตาและจมูกของฮีโร่ได้อย่างถูกต้อง แต่ส่วนปากจะถูกลบ และโมเดลอื่นๆ สามารถสร้างภาพได้เพียงครึ่งส่วนแต่ยังมีความสอดคล้องกับลักษณะตัวละคร โดยรูปภาพที่โมเดลสร้างออกมาจากทั้ง 3 โมเดล มีสีที่ค่อนข้างเข้มและสด ซึ่งมีลักษณะเหมือนกับรูปภาพของฮีโร่เกม DOTA2 อีกด้วย ด้านผลลัพธ์เมตริก Frechet Inception Distance (FID) ภาพที่สร้างขึ้นที่มีคะแนน FID ต่ำมีความใกล้เคียงในฐานภาพสไตล์ DOTA2 ดังเดิม ในเชิงคุณภาพพบว่ากลุ่มตัวอย่างชื่นชอบ โมเดล1 มากที่สุด ตามด้วยโมเดล2 และ โมเดล3 ตามลำดับ และปัจจัยหลักที่ตัดสินใจเลือกมาจากความมีกลิ่นอายของ DOTA2 เป็นสำคัญ จึงสรุปได้ว่างานวิจัยนี้สามารถเป็นแนวทางเพื่อสร้างภาพตัวละครใหม่จากชุดตัวละครฮีโร่เกม DOTA2 เดิม มีโอกาสไปพัฒนาสร้างรายได้ให้กับบริษัทผู้พัฒนาเกม ความพึงพอใจของผู้เล่น และองค์ความรู้ในการพัฒนางานเกมต่อไปโดยใช้เทคนิค GAN ได้

คำสำคัญ: Generative Adversarial Networks (GAN), Deep Learning, DOTA2 , StyleGAN, Few-shot Image Generation

Abstract

DOTA2 is a popular game in the E-sport game industry with a high market value. In particular, if developing a variety of in-game character traits, this can increase the company's revenue and player satisfaction. Many studies can demonstrate the effectiveness of Generative Adversarial Networks (GAN) applications to create cartoon images or recreate portraits. Thus, this research aims to study the effectiveness of the Style-GAN application to be used in recreating and redesigning the DOTA2 character. The form of creating a small number of shots through a few-shot Image Generation via Cross-domain Correspondence, which has the distinction of not requiring the large number of images used in training as a regular GAN and not overfitting with too much accuracy. The dataset used in training was the thumbnail of the DOTA2 hero character, with a size of 256x144 pixels, 111 images, and made up of 3 models: 1) the Source_ffhq model (a styleGAN2 model that learned to create human faces), 2) the Church models (styleGAN2 models that learned to create church images), and 3) the Horse models (a styleGAN2 model that learned to create horses). Quantitative comparisons showed that models adapted from the FFHQ model could correctly adapt a person's eyes and nose to the character's eyes and nose. However, the mouth part was erased, and other models could only create half the image but still correspond to the character traits. The images created by the three models were quite dark and fresh, which also looked like the images of a DOTA2 hero. In the Fréchet Inception Distance (FID) metric results, images created with a low FID score were as close as the original DOTA2-style images. As for qualitative evaluation, it was found that the sample of players have preferred the Source_ffhq model the most, followed by Church models and Horses models, respectively. The main factors were decided by being familiar with DOTA2. Therefore, this research can guide the creation of new character images from the original DOTA2 hero character series, with the opportunity to increase revenue for the game developer company, player satisfaction, and knowledge of further game development using GAN techniques.

Keywords: Generative Adversarial Networks (GAN), Deep Learning, DOTA2 , StyleGAN, Few-shot Image Generation

บทนำ

ในปัจจุบันอุตสาหกรรมเกมและ E-Sport กำลังเติบโตขึ้นอย่างรวดเร็วทั่วทุกมุมโลก อันเนื่องมาจากจุดเด่นหลายประการ เช่น ความสนุกสนาน ความตื่นเต้นท้าทาย มีกฎ กติกา ชัดเจนเปรียบเป็นกีฬาชนิดหนึ่ง และสามารถจัดการแข่งขันขึ้นที่ใดก็ได้บนโลก เป็นต้น Grand View Research (2020) ได้จัดทำรายงานวิเคราะห์ตลาดพบว่า ในปี 2020 อุตสาหกรรมเกมและ E-Sport มีส่วนแบ่งการตลาดอยู่ที่ 1.48 พันล้านเหรียญสหรัฐ และมีแนวโน้มจะเพิ่มสูงขึ้นเรื่อยๆ คาดการณ์อยู่ที่ 6.81 พันล้านเหรียญสหรัฐ ในปี 2027

E-Sport Game ที่ได้รับความนิยมสูง ได้แก่ เกม DOTA2 ซึ่งบริษัท วาล์วคอร์ปอเรชันได้พัฒนาขึ้น มีผู้เล่นมากกว่า 10 ล้านคนทั่วโลก (Clement, 2021) และจัดการแข่งขันในระดับภูมิภาคและระดับโลกอย่างต่อเนื่อง เพราะผู้เล่นสามารถเลือกใช้งานตัวละครได้เป็นจำนวนมากถึง 121 ตัวละคร โดยตัวละครจะถูกจำแนกออกเป็น 3 กลุ่ม ได้แก่ 1. ตัวละคร Strength จะมีภาพลักษณ์ที่แข็งแรง ดุดัน น่าเกรงขาม มีพลังทางกายภาพสูง 2. ตัวละคร Agility จะมีภาพลักษณ์ที่คล่องแคล่ว ว่องไว ปราดเปรียว 3. ตัวละคร Intelligence จะมีภาพลักษณ์ที่บอบบาง อ่อนแอ ทรงความรู้ เป็นผู้ใช้เวทมนตร์ เป็นต้น

จากการทบทวนวรรณกรรมพบว่าการสร้างตัวละครใหม่เพิ่มเติมขึ้นมาจะส่งผลให้ผู้เล่นมีทางเลือกในการเล่นมากขึ้น และหากตัวละครใหม่นั้นถูกออกแบบมาอย่างสวยงาม มีกลิ่นไอความเป็นตัวละครของ DOTA2 ตามประเภทของตัวละครที่กล่าวมาข้างต้นอย่างถูกต้อง จะส่งผลให้ผู้เล่นตัดสินใจจ่ายเงินซื้อตัวละครนั้นมาใช้งานมากขึ้น เนื่องจากเอกลักษณ์ของตัวละครมีความสัมพันธ์เชิงบวกกับความตั้งใจที่จะซื้ออุปกรณ์และสิ่งต่างๆ ในเกม (Park & Lee, 2011) และความคาดหวังที่ตัวละครสามารถเติบโตเปลี่ยนแปลงไปดีขึ้นยังมีผลทางบวกต่อความภักดีของผู้เล่นเกมออนไลน์นั้นอีกด้วย (Alghifari & Halim, 2020) หากแต่การสร้างตัวละครใหม่ขึ้นมานั้นมีค่าใช้จ่ายในการออกแบบที่ค่อนข้างสูง อีกทั้งในบางครั้งอาจออกแบบมาได้ไม่ถูกใจผู้เล่นโดยรวม

อย่างไรก็ตาม เทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) สามารถประยุกต์ใช้ศิลปะินเอไอ หรือ Generative Adversarial Networks (GAN) ซึ่งเป็นโมเดลต้นกำเนิด (Generative Model) มาสร้างภาพเหมือน และภาพการ์ตูน (Krohn et al., 2020) ซึ่งจุดเริ่มการศึกษา Generative Adversarial Network (GAN) มาจากงานของ Goodfellow et al. (2014) ที่ศึกษาเรื่องคลาสของเฟรมเวิร์กการเรียนรู้ของเครื่องจักร ในลักษณะของโครงข่ายประสาทสองเส้นแข่งขันกันในเกม ในรูปแบบของผลรวมศูนย์ (a zero-sum game) โดยที่ตัวแทนด้านหนึ่งได้ผลงาน ส่วนตัวแทนอีกด้านสูญเสีย ดังนั้นคณะผู้วิจัยจึงเห็นว่าถ้า GAN มาทดลองเรียนรู้และสร้างตัวละคร DOTA2 ใหม่ โดยอ้างอิงความรู้ที่ได้จากการวิเคราะห์ตัวละครเดิม จะสามารถช่วยลดต้นทุนการสร้างตัวละครใหม่และส่งเสริมยอดขายโดยรวมได้เป็นอย่างดี

นอกจากนี้คณะผู้วิจัยยังได้ทบทวนวรรณกรรมงานวิจัยที่กล่าวถึง GAN สำหรับการสังเคราะห์ภาพสองมิติเป็นหลัก ซึ่งพบว่าเทคนิคที่ใช้ GAN แตกต่างกันสามารถให้ผลลัพธ์ที่ดีเยี่ยมเหมือนกัน ตัวอย่างเช่น Karras et al. (2021) ศึกษาการใช้โครงสร้าง Generative Adversarial Networks (GAN) ของ NVIDIA ที่ประยุกต์จาก Style-

Transfer เพื่อให้สามารถสกัด high-level attribute เช่น ท่าโพส ท่าทาง สเกล ของส่วนประกอบในภาพ ได้อย่างอัตโนมัติ แบบการเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) โดยใช้ชุดข้อมูล human faces (Flickr-Faces-HQ, FFHQ) ซึ่งมีความละเอียดสูง และ Style-GAN สำหรับการปรับสไตล์ของภาพในแต่ละ Convolutional Neural Network layer (ปรับค่าของ Latent Code) ซึ่งมีการใช้งานดังนี้ Mapping function(f) ที่ 8 เลขเออร์เชื่อมต่อกันอย่างสมบูรณ์ Synthesis network(g) 18 layers, มีการฝัง input latent code ไว้ใน intermediate latent space (W), มีการ Add gaussian noise เพื่อไม่ให้มีการ โกงเกิดขึ้นได้ และใช้ Adaptive instance normalization (AdaIN) เพื่อควบคุม Generator ทั้งหมด 26.2 M trainable parameters ผลลัพธ์ที่ได้มีความคมชัดและสัดส่วนของภาพที่ถูกต้องกว่า Traditional Generator นอกจากนี้ยังสามารถปรับ Style (latent) เพื่อผสมและจับคู่ สไตล์ของภาพ A และ B เข้าด้วยกันได้ด้วย

Hagiwara & Tanaka (2020) ทดลองใช้ Class Condition ใน Training Dataset ทำให้เสถียรเพื่อให้ได้ผลลัพธ์ที่มีคุณภาพสูงขึ้นซึ่งจำเป็นต้องมีปริมาณ Training Dataset เป็นอย่างมาก ในบทความนี้จะใช้ Class Condition กับ Training Dataset ที่น้อยจึงใช้เทคนิค Clustering และ Data Augmentation มาช่วย (เพิ่มปริมาณข้อมูลขึ้น 5 เท่า) โดยเริ่มจากการ Segmentation รูปภาพมาสคอตให้เหลือเป็นตัวละครเดี่ยวๆภายใต้ฉากหลังสีขาว และย่อให้เหลือเป็นภาพขนาด 144 x 128 พิกเซล จำนวน 4018 ภาพ จากนั้นทำ Data Augmentation ด้วย Image Data Generator class on Kera ได้รวม 20090 ภาพ ขั้นตอนถัดมานำภาพไปผ่าน Feature Extraction เพื่อให้ได้ Features ดังนี้ 1. RGB Feature: (ใช้ Traditional เช่น Histogram) แบ่งตามสีของมาสคอต จากนั้น apply X-means ทั้งหมด 10 ครั้ง และหาค่าเฉลี่ยจะได้ cluster จาก RGB 2. ResNet Feature: ใช้ ResNet-50 ในการทำ Feature Extraction และ apply K-mean++ จะได้ cluster จาก Resnet 3. Edge ResNet Feature: ใช้ Sobel เพื่อหา edge ของแต่ละภาพและ ใช้ ResNet-50 ในการทำ Feature Extraction กับ apply K-mean++ จึงได้ Clustering result จากนั้นนำไป embed กับค่า y ของแต่ละภาพ และสร้าง Generative Model โดยมีการใช้ ResBlock และ Self-Attention ทั้งในฝั่ง Generator และ Discriminator จากผลลัพธ์ที่ได้สรุปผลได้ว่าการใช้ Resnet Extraction จะได้ผลดีที่สุด ช่วยให้ result ที่ได้จาก GAN มีคุณภาพที่ดียิ่งขึ้น

Guruprasad et al. (2020) ทำการเทรนโมเดลเพื่อสร้างใบหน้าคนในรูปแบบภาพการ์ตูนโดยโมเดลใช้ GAN (Generative Adversarial Network) แบบพื้นฐาน (เนื่องจากผู้เขียนไม่ได้ลงรายละเอียดเกี่ยวกับโครงสร้างเพิ่มเติม) ผู้วิจัยเทรน GAN กับรูปภาพจากด้าเซต CartoonSet100K ซึ่งเป็นด้าเซตที่ประกอบไปด้วยใบหน้าคนในรูปแบบภาพการ์ตูนที่มีขนาดรูปภาพ 500x500 พิกเซล จำนวนหนึ่งแสนภาพ โดยผลลัพธ์จากการสร้างภาพใบหน้าคนจากโมเดลให้ผลลัพธ์ที่น่าพอใจ

Jin et al. (2017) ใช้ DRAGAN (Deep Regret Analytic Generative Adversarial Networks) ร่วมกับ การประยุกต์ใช้โครงสร้างโมเดล SRResNet เป็นส่วนประกอบใน Generator และ Discriminator ในการสร้างโมเดลเรียนรู้เพื่อสร้างหน้าตัวละครอนิเมะขึ้นมาใหม่ โดยข้อมูลรูปภาพส่วนใบหน้าตัวละครจากเกมส์ตั้งแต่ปี 2005 ที่ได้

จากเว็บไซต์จำนวน 31,255 รูปภาพ และมีการกำหนดเท็กลักษณะให้แก่รูปภาพตามลักษณะรูปภาพจำนวน 34 แท็ก เช่น แท็กผมสีแดง แท็กผมสีดำ แท็กหมวก เพื่อเป็นอินพุตเพิ่มเติมให้แก่ Generator ด้วย ในส่วน โมเดล Generator ใช้ 16 ResBlocks และใช้ 3 Sub-pixel CNN เป็น Feature Map Upscaler ในส่วน Discriminator ใช้ 10 ResBlocks นำ Batch Normalization Layer ออกไปทั้งหมด และมีการเพิ่ม Fully Connected Layer ต่อจาก Output ของ CNN Layer สุดท้าย เพื่อใช้ในการจำแนกแท็กของรูป ผลลัพธ์ของการสร้างภาพพบว่าโมเดลสามารถสร้างรูปภาพได้ดีในอนิเมะที่มีผมสีบลอนด์และตาสีฟ้า และทำได้ไม่ดีนักในอนิเมะที่สวมแว่นหรือมีการม้วนผม เนื่องจากเป็นคุณลักษณะที่พบได้น้อยในแคตตาล็อกที่ใช้เทรน ในการประเมินประสิทธิภาพของโมเดลผู้เขียนใช้ Precision เพื่อเปรียบเทียบระหว่างแท็กคุณลักษณะที่ได้จากภาพจริงเปรียบเทียบกับแท็กคุณลักษณะที่กำหนดพบว่าค่า Precision ในแต่ละแท็กให้ผลลัพธ์ดังนี้ 1.0 : ผมสีบลอนด์ ผมสีน้ำตาล ผมสีดำ ผมสีฟ้าวารี ผมขาว ผมสั้น หน้าแดง ตาสีฟ้า ตาสีแดง ตาสีน้ำตาล ตาสีเขียว ตาสีเหลือง ตาสีฟ้าวารี | 0.95: ผมสีแดง หน้าลักษณะขี้ม หน้ามีการเปิดปาก ตาสีม่วง | 0.90: ผมสีเขียว | 0.85: ผมสีเงิน มีริบบิ้น ตาสีส้ม | 0.80: ผมสีชมพู ตาสีดำ | 0.75: ผมสีม่วง | 0.70: ผมสีฟ้า | 0.65: ผมสีส้ม | 0.60: ผมสีขาว ผมเปียคู่ ตาสีชมพู | 0.45: ผมหางม้า สวมแว่น | 0.35: ผมสีเทา | 0.20: ผมม้วน | 0.15: สวมหมวก

Tseng et al. (2021) งานนี้นำเสนอวิธีการฝึกสอนให้เป็นมาตรฐาน สำหรับ GAN ภายใต้การตั้งค่าข้อมูลที่จำกัด เพื่อบรรลุวัตถุประสงค์การฝึกอบรมที่แข็งแกร่งยิ่งขึ้นโดยใช้ชุดข้อมูล CIFAR 10/100 และ ImageNet ในการทดลอง ทำการทดลองในการสร้างภาพต่างๆ ชุดข้อมูลที่มีแกนหลัก GAN ต่างกันเพื่อสาธิตประสิทธิภาพของโครงการที่เสนอว่า 1) ปรับปรุงประสิทธิภาพของรุ่น GAN โดยเฉพาะภายใต้ข้อจำกัดการตั้งค่าข้อมูล และ 2) สามารถนำไปใช้กับ วิธีการเสริมข้อมูล (data augmentation methods) เพื่อเพิ่มประสิทธิภาพการทำงานต่อไป ในอนาคต นักวิจัยวางแผนปัญหาการขาดแคลนข้อมูลการฝึกอบรมสำหรับ 1) งาน GAN แบบมีเงื่อนไข เช่น image extrapolation, image to image translation ฯลฯ และ 2) การเรียนรู้ GAN ที่แข็งแกร่ง เกี่ยวกับข้อมูล large-scale noisy training

Back (2021) อภิปรายว่าเนื่องจากนักวิจัยหลายคนใช้โมเดลกำเนิดตามสไตล์ (The style-based generative model) ทำงานเช่นการแก้ไขภาพและการแปรภาพโดยสังเกตถึงความทรงพลังของสไตล์ตามสถาปัตยกรรม ซึ่งการแปรรูปภาพเป็นรูปภาพ (I2I) แบบไม่มีผู้ดูแล (Unsupervised image) แต่อย่างไรก็ตามความไม่สมดุลของข้อมูล การเรียนรู้การแจกแจงร่วมกันสำหรับโดเมนที่หลากหลายยังคงเป็นงานที่ยาก โมเดลที่มีอยู่สามารถสร้างภาพเป้าหมายที่เหมือนจริงได้ แต่การรักษาโครงสร้างของภาพต้นฉบับนั้นเป็นเรื่องยาก อีกทั้งการเทรน generative model ที่มีชุดข้อมูลขนาดใหญ่ในโดเมนต่างๆ ยังใช้เวลาและพลังคอมพิวเตอร์ในการประมวลผลเป็นจำนวนมาก ดังนั้นผู้วิจัยเสนอวิธี Cartoon-StyleGAN: Fine-tuning StyleGAN2 for Cartoon Face Generation แบบ novel unsupervised Image-to-Image translation methods ซึ่งวิธีการสำหรับการแปรรูปภาพเป็นรูปภาพ (I2I) แบบไม่มีผู้ดูแลโดยใช้การเรียนรู้การถ่ายโอนไปยัง โมเดลล่วงหน้าของ StyleGAN2 และยังเสนอวิธีการใหม่เพื่อรักษาโครงสร้างของภาพต้นฉบับและสร้างภาพที่เหมือนจริงในโดเมนเป้าหมาย โดยใช้ข้อมูล Flickr-Faces-HQ (FFHQ) ในส่วนของการฝึกของผู้วิจัย

ใช้การสูญเสียโครงสร้างกับสถาปัตยกรรมการข้ามอินพุต/เอาต์พุตของโมเดล StyleGAN2 ใช้การสูญเสียนี้สำหรับเลเยอร์ความละเอียดต่ำสามชั้นของตัวสร้างแหล่งที่มาและตัวสร้างเป้าหมาย และเรานำ $mse-loss$ มาใช้ เพื่อเป็นฟังก์ชันการสูญเสีย นอกจากนี้ การสูญเสียโครงสร้างยังช่วยให้สามารถสลับเลเยอร์ได้อย่างมีประสิทธิภาพ แม้ในเลเยอร์ที่มีความละเอียดต่ำ ผลการทดลองแสดงให้เห็นว่าวิธีการเหล่านี้มีประสิทธิภาพในการทำรูปภาพต้นฉบับและรูปภาพเป้าหมายคล้ายกันและช่วยสร้างภาพที่สมจริงยิ่งขึ้น อย่างไรก็ตาม การศึกษานี้จำเป็นต้องปรับเลเยอร์เพื่อเพิ่มประสิทธิภาพให้กับชุดข้อมูลแต่ละชุด ในการทำงานในอนาคตคาดว่าจะมีเสถียรภาพและเพิ่มประสิทธิภาพมากขึ้น

Li et al. (2021) นำเสนอ novel framework ที่สามารถสังเคราะห์รูปถ่ายหน้าคนให้หน้าตัวการ์ตูนอนิเมะซึ่งแตกต่างจากการแปลงในแบบอื่นๆที่มีก่อนหน้านี้โดยผู้ใช้รูปภาพอ้างอิงจากรูปภาพอนิเมะเข้ามาช่วยสังเคราะห์ด้วย ผู้แต่งนำเสนอ novel GAN-based anime-face translator (AniGAN) ที่สามารถสังเคราะห์หน้าตาตัวละครอนิเมะคุณภาพสูงได้ ซึ่งมีสถาปัตยกรรม generator ที่ถ่ายโอนรูปแบบ สี พื้นผิว ให้กลายเป็นโฉมหน้าอนิเมะพร้อมๆกัน โดยสถาปัตยกรรมมีการใช้ encoder แยกระหว่างรูปภาพพื้นผิวหน้าคนกับภาพสไตล์อนิเมะ และ decoder ที่แปลงภาพทั้งหลายให้กลายเป็นรูปภาพสไตล์อนิเมะที่สะท้อนจากรูปภาพหน้าคน โดย encoder ประกอบไปด้วยตัว Adaptive Stack convolutional (ASC) block ซึ่งเป็นการผสมกันอย่างง่ายของ convolutional layer กับ Adaptive Instance Normalization (AdaIN) normalization layer ที่ปราศจาก residual connection และ Fine-grained Style Transfer (FST) block ที่สามารถถ่ายทอด สี และ รายละเอียดของภาพโดยการผสมของ upsampling, convolutional, และ AdaIN normalization layers ในส่วนของ discriminator เป็นสถาปัตยกรรม double-branch discriminator เนื่องจากด้วยใบหน้าของอนิเมะและหน้าของคนมีการกระจายตัวที่สร้างข้อมูลบนใบหน้ามนุษย์เหมือนกัน ดังนั้น discriminator ช่วยลดการบิดเบือนและคงรูปแบบใบหน้าอนิเมะโดนประกอบด้วย 2 แขนงนั่นคือ 1. การแยกแยะภาพจริงและภาพไม่จริงของใบหน้าอนิเมะ 2. การแยกแยะภาพจริงและไม่จริงของใบหน้ามนุษย์ ซึ่งจะเกิดการเรียนรู้และปรับให้ได้อย่างมีประสิทธิภาพ ชุดข้อมูลที่ใช้ได้แก่ Selfie2anime และ Face2anime จากผลการทดลองสามารถสรุปได้ว่าวิธีของผู้วิจัยได้รับผลลัพธ์ที่ดีกว่าเทียบกับวิธีการ state-of-the-art ต่างๆ

Vavilala & Forsyth (2022) อภิปรายว่าเนื่องจาก Style-Gan ถูกใช้อย่างหลากหลายในงานปรับแต่งรูปภาพต่างๆรวมทั้งในงานการ์ดด้วยแต่ว่าการปรับแต่งนี้ยังไม่ส่งผลให้ความแตกต่างจากภาพต้นฉบับและภาพที่ถูกสร้างอย่างมีนัยยะสำคัญได้เลย ดังนั้นผู้เขียนจึงปรับแต่ง Style-Gan ทำให้การปรับเปลี่ยนมีความแตกต่างมากยิ่งขึ้น ชุดข้อมูลที่เข้ามาจาก the Yu-Gi-Oh card art dataset มี 11k ตัวอย่างในสไตล์และคลาสที่ต่างกันออกไป เริ่มจากการทำความสะอาดข้อมูลโดยใช้สายตาคัดแยกจากมนุษย์และคัดแยกออกประมาณ 500 ตัวอย่างเนื่องจากไม่สามารถระบุการ์ดนั้นได้และทำการควมแซมเปิดข้อมูลให้เป็น 256-res และเวอร์ชัน 512-res ทั้งชุดข้อมูล ผู้เขียนได้ทำการฝึก 512-res สำหรับการ์ดประเภทมอนสเตอร์ด้วย batch size ขนาด 96 และ 227 ชั่วโมงในการฝึกสำหรับรูปภาพ 25 ล้านภาพจากการใช้ adaptive discriminator augmentations (ADA) โดยที่ generator และ discriminator มีพารามิเตอร์ 28.7 ล้าน และ 28.9 ล้าน ตามลำดับ และใช้ hyperparameter สำหรับ R1 regularization ซึ่งการรันที่ดีที่สุดได้ค่า FID ที่

10.73 มีค่าน้อยซึ่งยิ่งดี จากการทดลองสามารถสรุปผลได้ว่า Style-GAN2 สามารถสร้างการ์ดที่น่าดึงดูดได้และมีความแตกต่างอย่างมีนัยยะสำคัญได้

Ojha et al. (2021) ได้ศึกษาการสร้าง GAN ในลักษณะ การสร้างภาพข้อความน้อยผ่านการโต้ตอบข้ามโดเมน (Few-shot Image Generation via Cross-domain Correspondence) เนื่องจากผู้วิจัยเห็นว่าการเทรนโมเดล GANs ซึ่งโดเมนเป้าหมายมีตัวอย่างจำกัด สามารถส่งผลให้เกิดโอเวอร์ฟิตได้อย่างง่ายดาย จึงพยายามที่จะใช้ประโยชน์จากโดเมนต้นทางขนาดใหญ่สำหรับการเทรนล่วงหน้า (pretraining) และถ่ายโอนความหลากหลายข้อมูลจากต้นทางสู่เป้าหมาย ในลักษณะให้รักษาความเหมือนและความแตกต่างที่สัมพันธ์กันระหว่างอินสแตนซ์ในแหล่งที่มา โดยใช้ระยะห่างของการสูญเสียคงที่ในโดเมนข้ามแบบใหม่ (novel cross-domain distance consistency loss) และเพื่อลดการเกิดโอเวอร์ฟิตเกินไป จึงใช้กลยุทธ์ที่ยึดตามจุดยึด (anchor-based strategy) เพื่อส่งเสริมระดับความแตกต่างของความสมจริงเหนืออาณาเขตต่าง ๆ ในพื้นที่แฝง จากการทดลองนี้ทั้งผลการทดลองเชิงคุณภาพและเชิงปริมาณ ผู้วิจัยได้พบว่า การสร้างภาพโมเดลข้อความน้อยจะค้นหาคำตอบโดยอัตโนมัติระหว่างโดเมนต้นทางและปลายทาง และสร้างภาพที่มีความหลากหลายและสมจริงมากกว่าวิธีการก่อนหน้านี้

จากการทบทวนวรรณกรรมที่กล่าวมาข้างต้น สรุปโดยภาพรวมจากงานวิจัยสามารถแสดงถึงประสิทธิภาพของการประยุกต์ใช้ GAN ได้เป็นอย่างดี กล่าวคือ Karras et al. (2021) เสนอการผสมสไตล์ของภาพวัตถุประเภทเดียวกันเข้าด้วยกันเช่น หน้าคน, แมว และ รถยนต์ เป็นต้น โดยการปรับแต่ง generator จากพื้นฐานของ Style-GAN Hagiwara & Tanaka (2020) เสนอการใช้เทคนิค feature extraction เพื่อหา feature ที่ทำ clustering result เพื่อทำ generative model ที่ดีที่สุดสำหรับการสร้างมาสคอต Jin et al. (2017) เสนอการปรับแต่ง generator และ discriminator เพื่อให้สร้างภาพการ์ตูนอนิเมะจากภาพเกมอนิเมะที่รวบรวมมาอย่างมีประสิทธิภาพ Guruprasad et al. (2020) และ Tseng et al. (2021) เสนอการใช้ GAN ในรูปแบบพื้นฐานและเสนอการเทรน GAN อย่างเป็นมาตรฐานตามลำดับ Back (2021) และ Li et al. (2021) เสนอการสังเคราะห์รูปภาพโดยเน้นการคงสภาพโครงสร้างของรูปภาพเดิมเอาไว้พร้อมกับการเปลี่ยนเป็นภาพในรูปแบบการ์ตูนหรืออนิเมะ ในขณะที่ Vavilala & Forsyth (2022) เสนอการสร้าง ความแตกต่างอย่างมีนัยยะสำคัญระหว่างภาพต้นตำหรับและภาพที่ถูกสร้างสำหรับการสร้างภาพการ์ดที่แตกต่างกัน อย่างไรก็ตามกรณีที่โดเมนเป้าหมายมีตัวอย่างจำกัด สามารถส่งผลให้เกิดโอเวอร์ฟิตได้อย่างง่ายดาย Ojha et al. (2021) ได้เสนอการสร้าง GAN ในลักษณะ การสร้างภาพข้อความน้อยผ่านการโต้ตอบข้ามโดเมน เพื่อใช้ประโยชน์จากโดเมนต้นทางขนาดใหญ่ในการเทรนล่วงหน้า (pretraining) และถ่ายโอนความหลากหลายข้อมูลจากต้นทางสู่เป้าหมาย ทำให้ลดปัญหาที่กล่าวมาได้

ดังนั้นแสดงให้เห็นว่า ในปัจจุบันนักวิจัยได้นำ Style-Gan มาใช้ในการสร้างภาพการ์ตูน ภาพมาสคอต หรือสร้างภาพบุคคลขึ้นใหม่อย่างแพร่หลายทาง คณะผู้วิจัยจึงเล็งเห็นความเป็นไปได้ในการประยุกต์ใช้ Style-Gan เพื่อใช้ในการสร้างและออกแบบตัวละคร DOTA2 ขึ้นมาใหม่ โดยจะนำ Style-Gan Model ที่ถูกเทรนไว้ดีแล้วมาปรับปรุงและเทรนเพิ่มเติมจากภาพตัวละคร DOTA2 เพื่อให้สามารถสร้างภาพตัวละครที่มีเอกลักษณ์ตรงตามความ

ต้องการได้ต่อไป โดยคณะผู้วิจัยจะป้องกันการเกิดโอเวอร์ฟิตในการศึกษาครั้งนี้ เนื่องจากข้อมูลตัวอย่างที่นำมาเทรนมีจำนวนน้อย จึงจะใช้แนวทางของ Ojha et al. (2021) ในลักษณะการสร้างภาพข้อความน้อยผ่านการโต้ตอบข้ามโดเมน (Few-shot Image Generation via Cross-domain Correspondence) โดยพยายามที่จะใช้ประโยชน์จากโดเมนต้นทางขนาดใหญ่สำหรับการเทรนล่วงหน้า (pretraining) และถ่ายโอนความหลากหลายข้อมูลจากต้นทางสู่เป้าหมายอีกด้วย

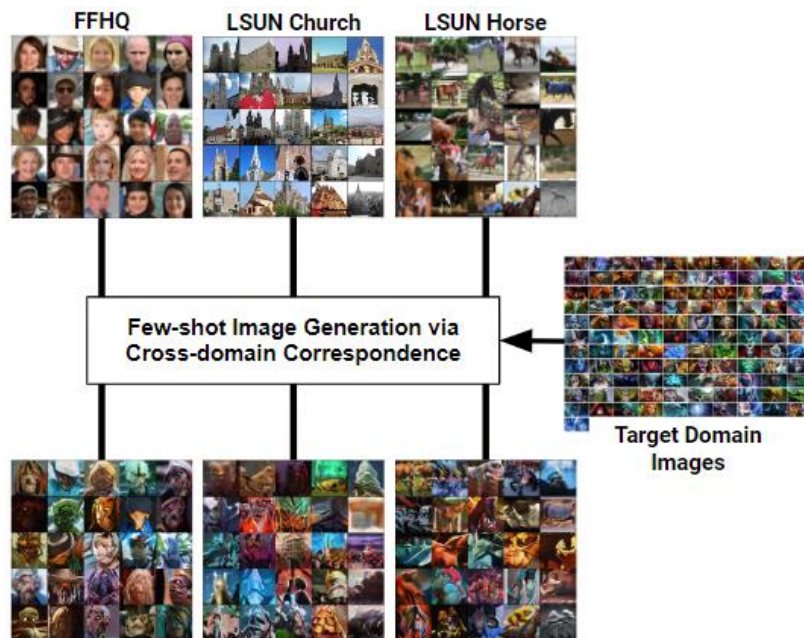
วิธีการทดลอง

Few-shot image generation

งานวิจัยนี้ประยุกต์ใช้เทคนิคการเรียนรู้จากงานวิจัย Few-shot Image Generation via Cross-domain Correspondence (Ojha et al., 2021) ซึ่งเป็นเทคนิคที่ทำให้ Pre-trained Generator สามารถประยุกต์เพื่อสร้างภาพออกมาในรูปแบบที่ผู้ใช้งานต้องการได้ โดยมีจุดเด่นคือ ใช้จำนวนรูปภาพขั้นต่ำสำหรับการเทรนเพียง 1 ภาพ โดยอ้างอิงจากตัวอย่างการทดลองในงานวิจัย Few-shot Image Generation via Cross-domain Correspondence ที่มีการใช้รูปภาพสำหรับเรียนรู้เพียง 1 ภาพ แต่ก็สามารถสร้างผลลัพธ์ออกมาได้อย่างน่าพอใจ แต่จำนวนรูปภาพสำหรับการเรียนรู้ที่เพิ่มขึ้น ก็ทำให้ผลลัพธ์ดีขึ้นหรือมีความหลากหลายมากขึ้นเช่นเดียวกัน (Ojha et al., 2021) แตกต่างจากการเทรน GAN (Generative Adversarial Networks) ที่จำเป็นต้องการรูปภาพสำหรับการเทรนจำนวนมาก โดยงานวิจัย Few-shot Image Generation via Cross-domain Correspondence (Ojha et al., 2021) ใช้เทคนิค Cross-domain Correspondence ที่มีการใช้ Regularization เพื่อให้โมเดลสามารถส่งผ่านความแตกต่างที่เกิดขึ้นใน Feature ของ Source Domain ไปยัง Domain Destination ได้ (พยายามทำให้ค่าความแตกต่างระหว่าง Feature ใน Target Domain มีค่าใกล้เคียงกันกับค่าความแตกต่างของ Feature ที่จะเกิดขึ้นใน Source Domain) นอกจากนี้เทคนิคนี้จะพยายามทำให้ Feature Distribution ของ Target Domain มีลักษณะใกล้เคียงกับ Feature Distribution ของ Source Domain อีกด้วย โดยเทคนิคนี้ต้องการอินพุต 2 ประเภท ได้แก่ Pre-trained Model ในรูปโครงสร้าง styleGANv2 และภาพ Target Domain อย่างน้อย 1 ภาพ ซึ่งภาพ Target Domain จะเป็นภาพที่ต้องการให้ Pre-trained Model ทำการเรียนรู้และประยุกต์โมเดลไปสร้างภาพในรูปแบบของ Target Domain ที่กำหนด หลังจากการเทรน ผลลัพธ์ที่ได้คือ Generator ที่สามารถสร้างรูปภาพในลักษณะเดียวกันกับ Pre-trained Model แต่มีการประยุกต์ Domain ที่ต้องการให้เรียนรู้เข้าไปด้วย ตัวอย่างเช่น กำหนด Pre-trained Model คือโมเดลสร้างภาพใบหน้าคน และกำหนดภาพ Target Domain คือรูปวาดใบหน้าคนของศิลปิน A ผลลัพธ์ที่ได้จากการเทรนโมเดลคือโมเดลสามารถสร้างรูปวาดใบหน้าคนในรูปแบบการวาดโดยศิลปิน A ได้ โดยรักษาโครงสร้างใบหน้าที่สำคัญจาก Pre-trained Model ได้โดยไม่เกิดการ Overfitting กับภาพ Target Domain ที่กำหนดให้เรียนรู้ แม้จะมีรูปภาพเพียง 1 ภาพ

เนื่องจากข้อจำกัดทางด้านจำนวน Training Image ซึ่งภาพของฮีโร่ในเกมส์ DOTA2 นั้นมีจำนวนไม่มากในการศึกษานี้เราจะหยิบยืมเทคนิคการเรียนรู้จากงานวิจัย Few-shot Image Generation via Cross-domain

Correspondence (Ojha et al., 2021) มาใช้กับ Pre-trained Model แบบต่างๆเพื่อใช้สร้างภาพอีโร DOTA2 ซึ่งงานวิจัยนี้จะทำการเทรนและเปรียบเทียบผลลัพธ์การสร้างภาพอีโรของเกมส์ DOTA2 จากผลลัพธ์ของ Pre-trained Model เริ่มต้นที่แตกต่างกัน 3 โมเดล (โมเดลรูปแบบ styleGANv2) ได้แก่ โมเดล FFHQ (Flickr-Faces-HQ) ที่ใช้สำหรับสร้างภาพใบหน้าคน, โมเดล LSUN (Large-scale Scene Understanding) Church ที่ใช้สำหรับสร้างภาพโบสถ์ และโมเดล LSUN Horse ที่ใช้สำหรับสร้างภาพม้า และกำหนดให้โมเดลเหล่านี้ทำการเรียนรู้การประยุกต์การสร้างรูปภาพโดยใช้รูปภาพอีโร DOTA2 จำนวน 111 ภาพ เป็นภาพ Target Domain และเปรียบเทียบผลลัพธ์ที่ได้จาก Generator หลังจากการเทรนแล้ว โดยโมเดลทุกประเภทจะกำหนดเงื่อนไขพารามิเตอร์ในการเรียนรู้ และรูปภาพ Target Domain ที่เหมือนกันทุกอย่าง แตกต่างกันเพียง Pre-trained Model ที่ไม่เหมือนกันเท่านั้น สรุปภาพโครงสร้างการทดลองในงานวิจัยนี้ ดังแสดงในภาพที่ 1 และกำหนดค่าและเงื่อนไขพารามิเตอร์ในการเรียนรู้ของโมเดลดังแสดงในตารางที่ 1



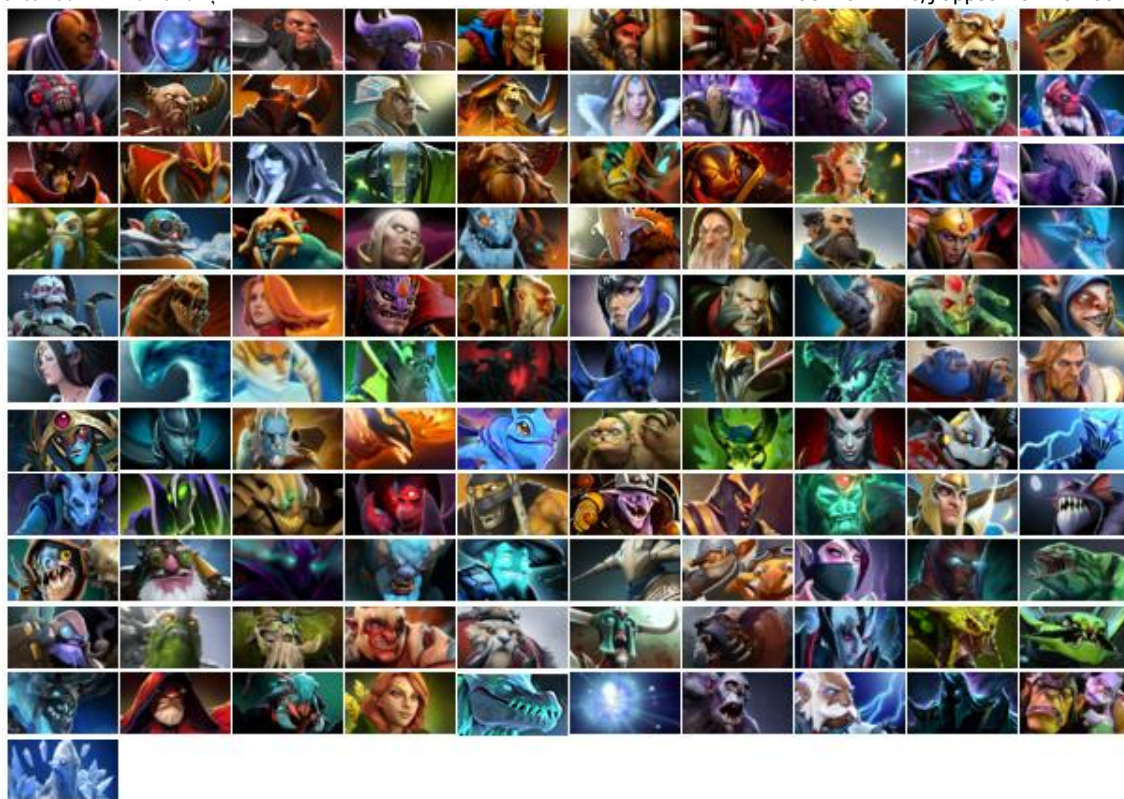
ภาพที่ 1 แสดงโครงสร้างการทดลองในงานวิจัย

Parameter	Value
Iteration	25000
Batch size	4
Generated image size	256x256 px
Latent dimension	512
Learning rate	0.002
Augmentation	True

ตารางที่ 1 พารามิเตอร์ในการเรียนรู้ของโมเดล

ชุดข้อมูล(Dataset)

รูปใบหน้าตัวละครฮีโร่ของเกมส่ DOTA2 ขนาด 256x144 พิกเซล จำนวน 111 รูป (AucT, 2013) แสดงดังภาพที่ 2 จะถูกกำหนดให้เป็นภาพ Target Domain ให้โมเดลทำการเรียนรู้สำหรับทั้ง 3 Pre-trained Model โดยรูปภาพจำนวน 111 รูปถือว่าเป็นจำนวนที่มากพอแล้วสำหรับการใช้การประยุกต์การเรียนรู้ของ Pre-trained Model โดยใช้เทคนิค Cross-domain Correspondence อ้างอิงจากการทดลองในงานวิจัย Few-shot Image Generation via Cross-domain Correspondence ที่พบว่าจำนวนรูปภาพ Target Domain จำนวน 10 ภาพ ก็ให้ผลลัพธ์ที่น่าประทับใจแล้ว (Ojha et al., 2021)



ภาพที่ 2 รูปีฮีโร่เกมส์ DOTA2 ที่ใช้ในการเรียนรู้ จำนวน 111 รูป
ที่มา: (AucT, 2013)

Comparative Experiment

เหตุผลที่เปรียบเทียบผลลัพธ์จากโมเดลเริ่มต้นซึ่งสร้างภาพต่างชนิดกันอย่างชัดเจน คือ ภาพในหน้าคน (FFHQ) ภาพโบสถ์ (LSUN Church) และภาพม้า (LSUN Horse) เพราะต้องการเปรียบเทียบผลลัพธ์ที่ได้จากการประยุกต์โมเดลที่เป็นตัวแทนของส่วนประกอบหลักที่สำคัญในเกมที่ต่างกัน คือ FFHQ จะเป็นตัวแทนของคนหรือฮีโร่ในเกมส์, LSUN Church จะเป็นตัวแทนของสิ่งก่อสร้างในเกมส์ เช่น บ้าน, ร้านค้าหรือป้อมปราการ และ LSUN Horse จะเป็นตัวแทนของสัตว์ในเกมส์ เช่น สัตว์ที่ใช้ส่งไอเทมหรือสัตว์ที่เป็นยานพาหนะของฮีโร่บางประเภท ว่าจะมีผลลัพธ์เป็นอย่างไร

การทดลองในงานวิจัยนี้ทดลองโดยใช้ Google Colab Pro (CPU: Intel® Xeon® CPU@2.00GHz, GPU: Tesla P100, RAM: 13GB) กำหนด โดยรายละเอียดรายละเอียดเวอร์ชันของไพทอน และ Library ที่ใช้ในงานวิจัยแสดงดังในตารางที่ 2














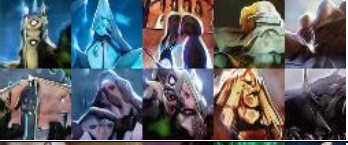













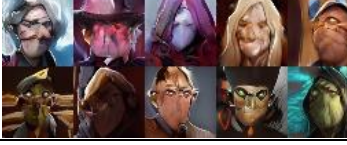








Program/Library	Version
Python	3.7.13
lpips	0.1.3
ninja	1.10.2.3
numpy	1.19.5
Pillow	8.0.1
pytorch-fid	0.1.1
torch	1.7.0
Torchaudio	0.7.0
torchvision	0.8.1
tqdm	4.51.0
visdom	0.1.8.9

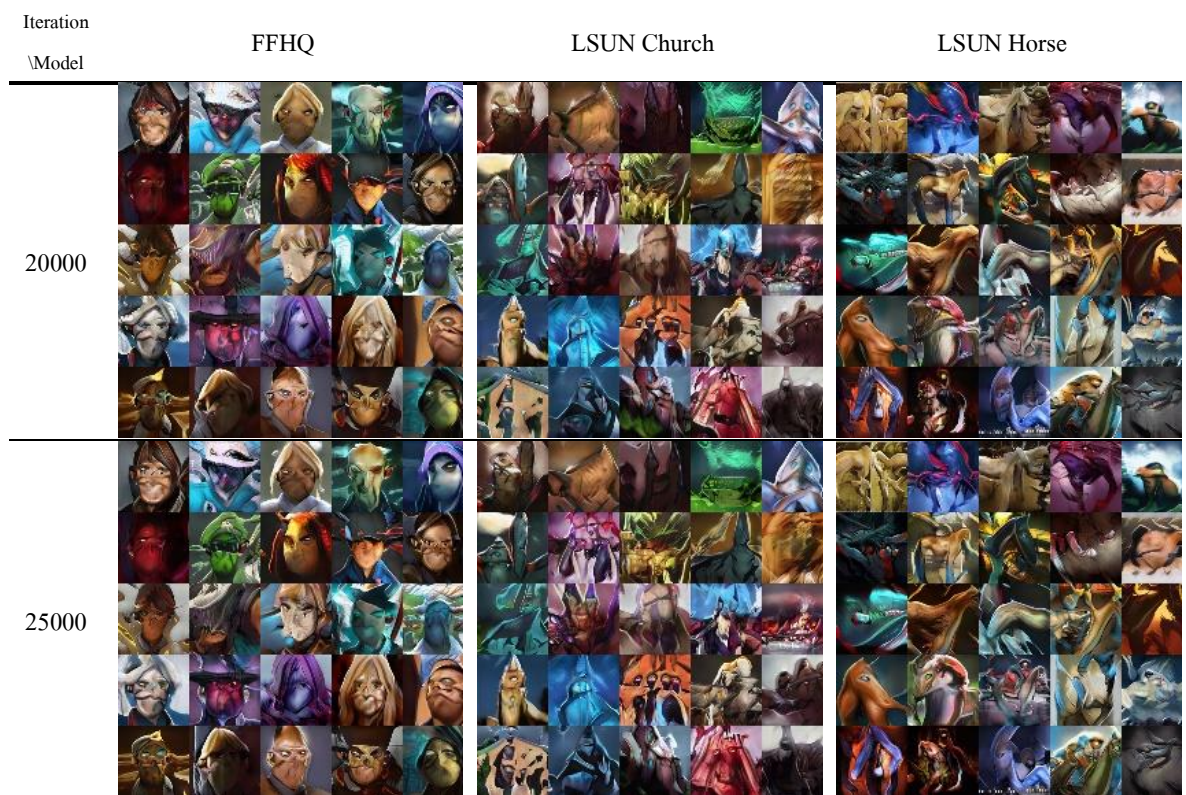
ตารางที่ 2 แสดงรายละเอียดเวอร์ชันของไพทอน และ Library ที่ใช้ในงานวิจัย

ผลการทดลอง

ในการศึกษานี้จะแสดงผลลัพธ์ที่ได้จากการสร้างภาพโดย Adapted Generator จาก Source Model ที่ต่างกัน ทั้ง 3 โมเดล (FFHQ, LSUN Church และ LSUN Horse) ตามที่ได้กล่าวไป โดยแสดงผลลัพธ์ทุกๆ 5000 Iteration ตั้งแต่ Iteration ที่ 0 (ยังไม่เกิดการ Adaptation) จนถึง Iteration ที่ 25000 โดยผลลัพธ์แสดงดังในตารางที่ 3

จากภาพผลการทดลองพบว่า ในส่วนของโมเดลที่ประยุกต์มาจากโมเดล FFHQ แสดงให้เห็นว่า โมเดลสามารถตัดแปลงส่วนตาและจมูกของคน มาเป็นตาและจมูกของฮีโร่ได้อย่างถูกต้อง แต่ส่วนปากจะถูกลบออกไป ในส่วนผลลัพธ์ที่ประยุกต์มาจากโมเดล LSUN Church แสดงให้เห็นว่าจำนวนภาพที่ถูกสร้างออกมาจะสามารถคงรูปร่างโบสถ์ไว้ได้เพียงครั้งหนึ่ง และยังพบได้ว่าโมเดลจะพยายามแทนที่หน้าต่างของโบสถ์ด้วยดวงตาของฮีโร่ และส่วนผลลัพธ์ที่ประยุกต์มาจากโมเดล LSUN Horse ก็มีผลลัพธ์ใกล้เคียงกันกับโมเดลจาก LSUN Church คือภาพผลลัพธ์ที่ถูกสร้างออกมาจะสามารถคงรูปร่างม้าไว้ได้เพียงครั้งหนึ่ง และจากการพิจารณาสีของรูปภาพที่โมเดลสร้างออกมาจากทั้ง 3 โมเดล จะมีสีที่ค่อนข้างเข้มและสด ซึ่งมีลักษณะเหมือนกับรูปภาพของฮีโร่เกมส์ DOTA2 ที่กำหนดให้โมเดลทำการเรียนรู้

Iteration	FFHQ	LSUN Church	LSUN Horse
Model			
0			
			
			
			
			
5000			
			
			
			
			
10000			
			
15000			



ตารางที่ 3 ผลลัพธ์ที่ได้จาก Adapted Generator จากโมเดล FFHQ, LSUN Church และ LSUN Horse ตั้งแต่ Iteration ที่ 0 ถึง 25000

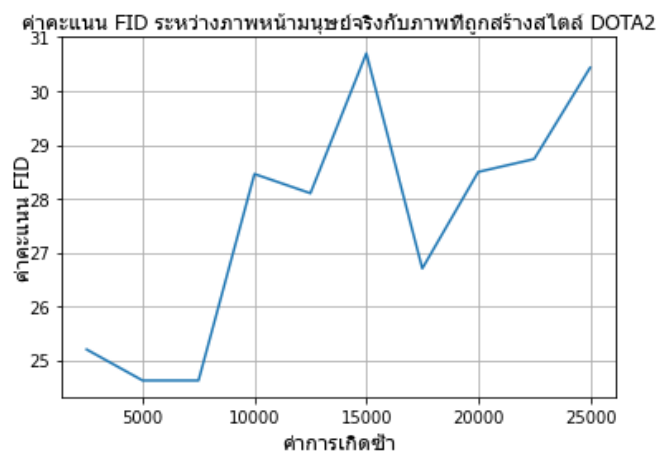
วิจารณ์ผลการทดลอง

การเปรียบเทียบเชิงปริมาณ (Quantitative comparison)

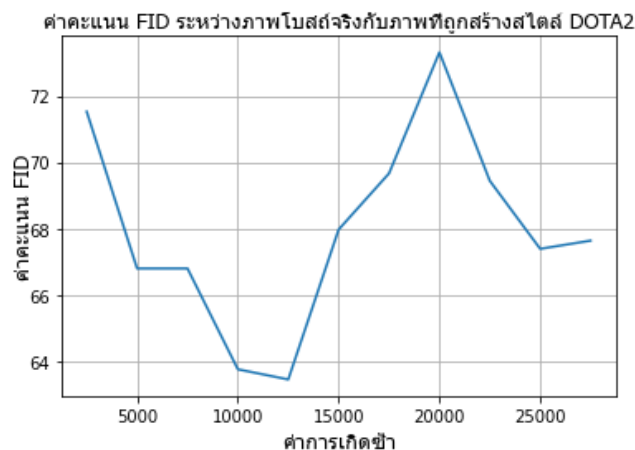
ในการเปรียบเทียบเชิงปริมาณ งานวิจัยนี้แสดงผลลัพธ์ที่ได้จากการสร้างภาพโดย Adapted Generator จาก Source Model ที่ต่างกันทั้ง 3 โมเดล (FFHQ, LSUN Church และ LSUN Horse) ตามที่ได้กล่าวไป โดยจากภาพผลการทดลองในตารางที่ 3 พบว่า ในส่วนของโมเดลที่ประยุกต์มาจากโมเดล FFHQ แสดงให้เห็นว่า โมเดลสามารถดัดแปลงส่วนตาและจมูกของคน มาเป็นตาและจมูกของฮีโร่ได้อย่างถูกต้อง แต่ส่วนปากจะถูกลบออกไป ในส่วนผลลัพธ์ที่ประยุกต์มาจากโมเดล LSUN Church แสดงให้เห็นว่าจำนวนภาพที่ถูกสร้างออกมาจะสามารถคงรูปร่างโบสถ์ไว้ได้เพียงครั้งหนึ่ง และยังพบได้ว่าโมเดลจะพยายามแทนที่หน้าต่างของโบสถ์ด้วยดวงตาของฮีโร่ และส่วนผลลัพธ์ที่ประยุกต์มาจากโมเดล LSUN Horse ก็มีผลลัพธ์ใกล้เคียงกันกับโมเดลจาก LSUN Church คือภาพผลลัพธ์ที่ถูกสร้างออกมาจะสามารถคงรูปร่างม้าไว้ได้เพียงครั้งหนึ่ง และจากการพิจารณาสีของรูปภาพที่โมเดลสร้างออกมาจากทั้ง 3 โมเดล จะมีสีที่ค่อนข้างเข้มและสด ซึ่งมีลักษณะเหมือนกับรูปภาพของฮีโร่เกมส์ DOTA2 ที่กำหนดให้โมเดลทำการเรียนรู้ จึงอภิปรายผลการทดลองได้ว่า คณะผู้วิจัยสามารถประยุกต์เทคนิคการเรียนรู้จากงานวิจัย Few-

shot Image Generation via Cross-domain Correspondence ตามแนวทางของ Ojha et al. (2021) ซึ่งเป็นเทคนิคที่ทำให้ Pre-trained Generator และไม่เกิด Overfitting มาสร้างภาพตัวละคร DOTA2 ได้

นอกจากนี้คณะผู้วิจัยได้ประเมินคุณภาพของผลลัพธ์ด้วยเมตริก Frechet Inception Distance (FID) ซึ่งเป็นที่นิยมใช้ในการประเมินคุณภาพของภาพสังเคราะห์ในงานสังเคราะห์ภาพคะแนน ซึ่ง FID จะประเมินความแตกต่างของการกระจายระหว่างใบหน้ามนุษย์จริง (Flickr-Faces-HQ (FFHQ)), ม้า (LSUN Horses), และภาพโบสถ์คริสตจักร (Church dataset) กับภาพที่สังเคราะห์ขึ้นมาสไลด์ DOTA2 คะแนน FID ที่ต่ำกว่าบ่งชี้ถึงความแตกต่างการกระจายภาพที่ถูกสร้างขึ้นนั้นคล้ายกับภาพสไลด์ DOTA2 ของจริงมากเพียงใด กล่าวคือภาพที่สร้างขึ้นที่มีคะแนน FID ต่ำมีความใกล้เคียงในฐานภาพสไลด์ DOTA2 ของจริง



ภาพที่ 3 ค่าคะแนน FID ระหว่างภาพหน้ามนุษย์จริงกับภาพที่ถูกสร้างสไลด์ DOTA2



ภาพที่ 4 ค่าคะแนน FID ระหว่างภาพโบสถ์จริงกับภาพที่ถูกสร้างสไลด์ DOTA2



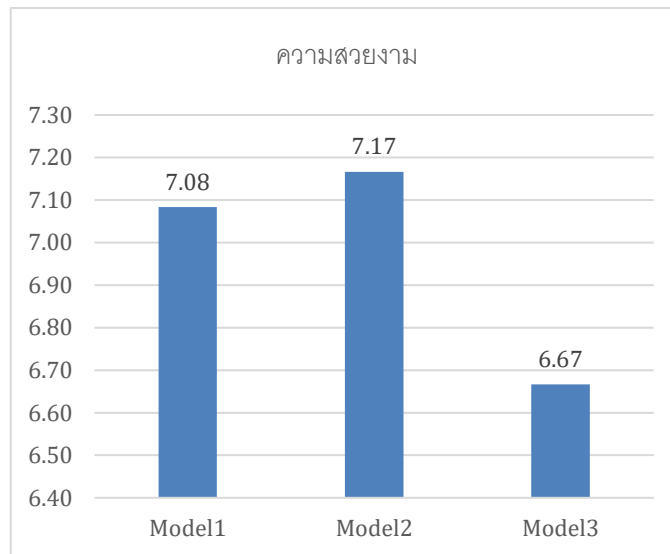
ภาพที่ 5 ค่าคะแนน FID ระหว่างภาพม้าจริงกับภาพที่ถูกสร้างสไตส์ DOTA2

จากภาพที่ 3-5 แสดงภาพแสดงค่าคะแนน FID ที่ขึ้นกับการเปลี่ยนไปของค่าการเกิดซ้ำ (iteration) จากผลการทดลองในเชิงปริมาณสามารถอธิบายได้ว่าค่าคะแนนของ FID ของทั้งสามโมเดล แต่ละโมเดลมีช่วงค่าการเกิดซ้ำ (iteration) ที่ต่ำที่สุดได้ดังนี้ โมเดล Source_ffhq มีช่วงค่าการเกิดซ้ำที่ 5000 โมเดล Church มีค่าการเกิดซ้ำที่ 12500 และ โมเดล Horses มีค่าการเกิดซ้ำที่ 7500

Qualitative comparison (การเปรียบเทียบเชิงคุณภาพ)

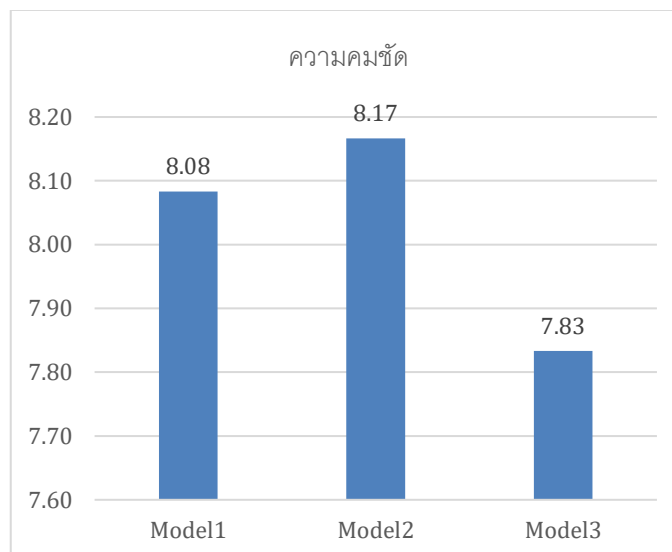
สำหรับการเปรียบเทียบเชิงคุณภาพ ทางคณะผู้วิจัยได้นำภาพที่ถูกสร้างขึ้นจากทั้ง 3 โมเดล คือ 1) โมเดล Source_ffhq (โมเดล styleGAN2 ที่เรียนรู้สร้างภาพใบหน้าคน), 2) โมเดล Church (โมเดล styleGAN2 ที่เรียนรู้สร้างภาพโบสถ์) และ 3) โมเดล Horses (โมเดล styleGAN2 ที่เรียนรู้สร้างภาพม้า) ไปให้ผู้เล่น DOTA2 ประเทศไทย ซึ่งมีประสบการณ์การเล่น DOTA2 1 ปีขึ้นไป และอยู่ในกลุ่ม DOTA2 Thailand จำนวนทั้งหมด 12 ท่าน เพื่อให้อ่านพิจารณาภาพที่ถูกสร้างขึ้นในหัวข้อต่อไปนี้

ด้านความสวยงาม พบว่า โมเดล2 ได้คะแนนเฉลี่ยสูงสุด 7.17 เต็ม 10 ตามด้วยโมเดล1 7.08 และ โมเดล3 6.67 คะแนนตามลำดับ ดังภาพที่ 6



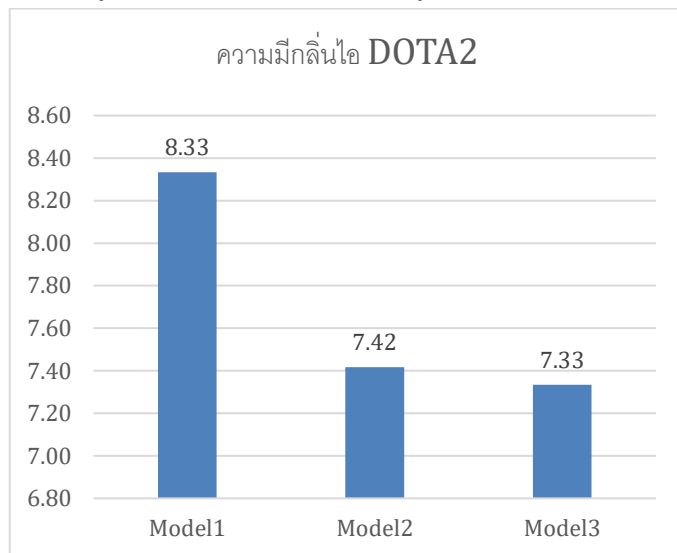
ภาพที่ 6 ความสวยงามของภาพที่ทดลองสร้างจากภาพตัวละคร DOTA2

ด้านความคมชัด พบว่า โมเดล2 ได้คะแนนเฉลี่ยสูงสุด 8.17 เต็ม 10 ตามด้วยโมเดล1 8.08 และ โมเดล3 7.83 คะแนนตามลำดับ ดังภาพที่ 7



ภาพที่ 7 ความคมชัดของภาพที่ทดลองสร้างจากภาพตัวละคร DOTA2

ด้านความมึนงง DOTA2 พบว่า โมเดล 1 ได้คะแนนเฉลี่ยสูงสุด 8.33 เต็ม 10 ตามด้วยโมเดล 2 7.42 และโมเดล 3 7.33 คะแนนตามลำดับ ดังภาพที่ 8 ทั้งนี้จากการสัมภาษณ์เพิ่มเติมพบว่าผู้เข้าร่วมการทดสอบส่วนใหญ่ลงความเห็นตรงกันว่าเมื่อได้เห็นภาพที่ถูกสร้างขึ้นจากโมเดล 1 สามารถรู้ได้ทันทีว่ามาจากเกม DOTA2



ภาพที่ 8 ความมึนงง DOTA2 ของภาพที่ทดลองสร้างจากภาพตัวละคร DOTA2

อย่างไรก็ตาม เมื่อพิจารณาไปในรายละเอียดจึงพบว่าเป็นตัวละครใหม่ที่ตนเองไม่รู้จัก ตัวอย่างเช่น ตัวละครใหม่สีแดงที่ถูกสร้างขึ้น ผู้เข้าร่วมการทดสอบ 7 จาก 12 ท่านลงความเห็นว่ามีแนวโน้มใกล้เคียงกับตัวละคร Grimstroke และ Axe ดังแสดงในภาพที่ 9



ตัวละครใหม่

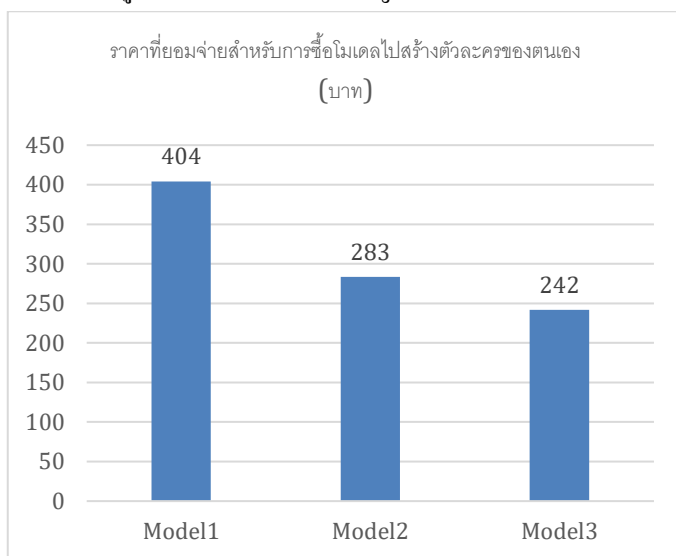
Grimstroke

Axe

ภาพที่ 9 ตัวอย่างภาพตัวละครที่ถูกสร้างขึ้นจากโมเดล 1 ที่มีความใกล้เคียงกับตัวละคร Grimstroke และ

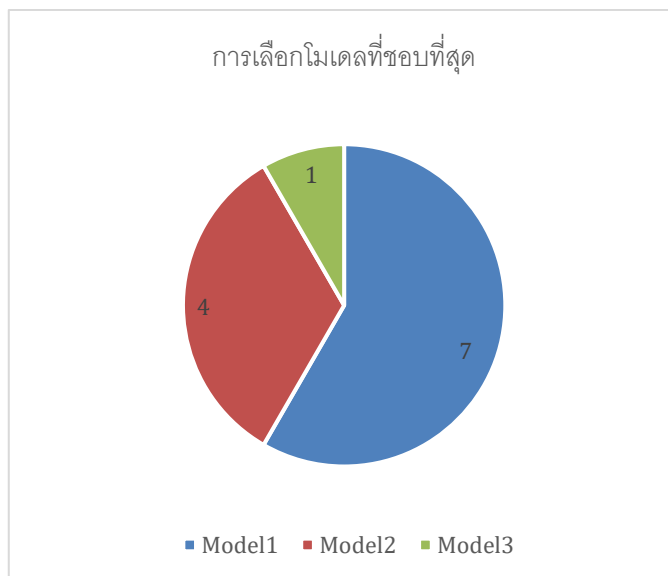
Axe

สำหรับราคาที่ยอมจ่ายสำหรับการซื้อโมเดลไปสร้างตัวละครของตนเอง พบว่า โมเดล1 มีราคาที่มีผู้เข้าร่วมการทดสอบยอมจ่ายเฉลี่ยสูงที่สุดรวม 404 บาท ตามด้วย โมเดล2 ราคา 283 บาท และ โมเดล3 ราคา 242 บาท ดังภาพที่ 10 ทั้งนี้จากการสอบถามเหตุผล ผู้เข้าร่วมการทดสอบต้องการนำโมเดลไปใช้สร้างภาพตัวละคร DOTA2 จากหน้าตัวเอง หรือใช้ไปสร้างภาพวาดจากผู้ชม (Fanart) เป็นส่วนใหญ่



ภาพที่ 10 ราคาที่ยอมจ่ายสำหรับการซื้อโมเดลไปสร้างตัวละครของตนเอง

ส่วนภาพรวมโมเดลที่ชอบที่สุด พบว่า โมเดล1 มีผู้เข้าร่วมการทดสอบชอบมากที่สุดทั้งหมด 7 คน ตามด้วยโมเดล2 4 คน และ โมเดล3 1 คนตามลำดับ โดยปัจจัยหลักที่ตัดสินใจเลือกมาจากความมีกลิ่นไอของ DOTA2 เป็นสำคัญ ดังภาพที่ 11



ภาพที่ 11 การเลือกโมเดลที่ชอบที่สุด

จากผลการทดลองในส่วนของการเสนอแนะ ผู้เข้าร่วมการทดสอบ มีความชื่นชอบและตื่นเต้นที่จะได้ทดลองสร้างตัวละคร DOTA2 ขึ้นมาใหม่ และมีข้อเสนอแนะเพิ่มเติมในรายละเอียดคือ การเพิ่มตัวเลือกในการระบุเพศ และ เผ่าพันธุ์ ของตัวละครที่จะสร้างขึ้น เช่น ตัวละครเพศหญิงเผ่าเอลฟ์ หรือ ตัวละครเพศชายเผ่าสัตว์ป่า เป็นต้น

ดังนั้นขณะนี้ผู้วิจัยจึงมีความเห็นว่าหากสามารถนำงานวิจัยทดลองสร้างภาพครั้งนี้ไปทำได้อีกจะมีประโยชน์มากในการสร้างตัวละครให้ได้ตามที่ต้องการมากยิ่งขึ้น รวมทั้งยังช่วยสร้างรายได้ให้กับบริษัทผู้พัฒนา เพิ่มความพึงพอใจของผู้เล่นเกม และเป็นงานต้นแบบที่ศึกษาการใช้เทคนิค GAN ในการพัฒนาเกมอีกด้วย ทั้งนี้เนื่องจากระยะเวลาทดลองจำกัด และอยู่ในช่วงการระบาดของโรคติดเชื้อไวรัส โควิด-19 ทำให้ในการประเมินเชิงคุณภาพมีจำนวนกลุ่มตัวอย่างเข้าร่วมน้อย ในการศึกษาในอนาคตจึงควรเพิ่มกลุ่มตัวอย่างให้มากขึ้น รวมถึงใช้เทคนิควิธีการวิจัยเชิงคุณภาพอื่นๆ เช่น การสัมภาษณ์เชิงลึก การอภิปรายกลุ่มย่อย เป็นต้น (Creswell & Creswell, 2018) เพื่อให้เกิดความเชื่อมั่นถึงผลลัพธ์ที่ได้อย่างมีประสิทธิภาพมากขึ้น

สรุปผลการทดลอง

งานวิจัยนี้ประยุกต์ใช้เทคนิคการสร้างภาพข้อความจำนวนน้อยผ่านการได้ตอบข้ามโดเมน (Few-shot Image Generation via Cross-domain Correspondence) โดยการใช้และเปรียบเทียบผลลัพธ์จากฟรีเทรนโมเดลที่แตกต่างกัน 3 โมเดล คือ 1) Flickr-Faces-HQ (FFHQ) 2) LSUN Horses และ 3) LSUN Horses เพื่อการเรียนรู้โดยใช้เทคนิค

การส่งผ่านการเรียนรู้จากภาพฮีโร่ของเกม Dota2 ส่วนข้อมูลที่ใช้ในการเรียนรู้ คือ รูปขนาดย่อของตัวละครฮีโร่ เกม Dota2 ขนาด 256x144 พิกเซล จำนวน 111 รูป โดยผลการเปรียบเทียบในเชิงปริมาณพบว่า โมเดลที่ประยุกต์มาจากโมเดล FFHQ แสดงให้เห็นว่า โมเดลสามารถดัดแปลงส่วนตาและจมูกของคน มาเป็นตาและจมูกของฮีโร่ได้อย่างถูกต้อง แต่ส่วนปากจะถูกลบออกไป ในส่วนผลลัพธ์ที่ประยุกต์มาจากโมเดล LSUN Church แสดงให้เห็นว่า จำนวนภาพที่ถูกสร้างออกมาจะสามารถคงรูปร่างโบสถ์ไว้ได้เพียงครั้งหนึ่ง และยังพบได้ว่าโมเดลจะพยายามแทนที่หน้าต่างของโบสถ์ด้วยดวงตาของฮีโร่ และส่วนผลลัพธ์ที่ประยุกต์มาจากโมเดล LSUN Horse ก็มีผลลัพธ์ใกล้เคียงกันกับโมเดลจาก LSUN Church คือภาพผลลัพธ์ที่ถูกสร้างออกมาจะสามารถคงรูปร่างม้าไว้ได้เพียงครั้งหนึ่ง และจากการพิจารณาสีของรูปภาพที่โมเดลสร้างออกมาจากทั้ง 3 โมเดล จะมีสีที่ค่อนข้างเข้มและสด ซึ่งมีลักษณะเหมือนกับรูปภาพของฮีโร่เกม Dota2 ที่กำหนดให้โมเดลทำการเรียนรู้ ในเชิงคุณภาพพบว่า โมเดล 1 เป็นโมเดลที่มีผู้เข้าร่วมการทดสอบชื่นชอบมากที่สุด ตามด้วยโมเดล 2 และ โมเดล 3 ตามลำดับ โดยปัจจัยหลักที่ตัดสินใจเลือกมาจากความมีกลิ่นอายของ Dota2 เป็นสำคัญ โดยผลการวิจัยส่วนใหญ่ยังสอดคล้องกับผลการทดลองของงานวิจัยที่มีมาก่อน ที่สามารถนำชุดภาพข้อมูลอื่นๆ มาใช้เทคนิค GAN เพื่อสร้างภาพใหม่อย่างเช่นชุดตัวละครฮีโร่เกม Dota2 และพัฒนาตัวละครในการสร้างรายได้ของบริษัทผู้พัฒนาเกม ความพึงพอใจของผู้เล่น และองค์ความรู้โดยใช้เทคนิค GAN ในการพัฒนาวงการเกมต่อไป

กิตติกรรมประกาศ

กลุ่มผู้วิจัยขอขอบพระคุณผู้เข้าร่วมการทดลองจากกลุ่ม Dota2 THAILAND ทั้ง 12 ท่านเป็นอย่างสูง ที่ได้สละเวลาเข้าร่วมทดสอบ แสดงความคิดเห็น และ ข้อเสนอแนะที่มีคุณค่าเพื่อปรับปรุงผลลัพธ์ของงานวิจัยให้ดียิ่งขึ้นไปในอนาคต และขอขอบพระคุณ ผศ.ดร.จิตรีศน์ สิริบรรณรัตนกุล ผู้เป็นที่ปรึกษา ให้คำแนะนำ ส่งผลให้งานวิจัยชิ้นนี้สำเร็จลุล่วงไปได้ด้วยดี

เอกสารอ้างอิง

- Alghifari, I. & Halim, R. (2020). Factors that influence expectancy for character growth in online games and their influence on online gamer loyalty. *Journal of Economics, Business, & Accountancy Ventura*, 22, 10.14414/jebav.v22i3.1873.
- AucT. (2013). Dota2 Hero Images. <https://auct.eu/dota2-hero-images/>
- Back J. (2021). Fine-Tuning StyleGAN2 For Cartoon Face Generation. *Computer Vision and Pattern Recognition*. <https://www.researchgate.net/publication/353069818>

Clement, J. (2021). *Number of monthly active users (MAU) of DOTA 2*

worldwide as of February 2021. Statista.

<https://www.statista.com/statistics/607472/DOTA2-users-number>

Creswell, J. W., & Creswell, J. D. (2018). *Research design: qualitative, quantitative, and mixed methods approaches*. 5th ed. Los Angeles: SAGE.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Nets*. Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.

Grand View Research (2020). *Esports Market Size, Share & Trends Analysis Report By Revenue Source (Sponsorship, Advertising, Merchandise & Tickets, Media Rights), By Region, And Segment Forecasts, 2020 - 2027*.

<https://www.grandviewresearch.com/industry-analysis/esports-market#>

Guruprasad, G., Gakhar, G., & Vanusha, D. (2020). *Cartoon character generation using generative Adversarial Network*. *International Journal of Recent Technology and Engineering (IJRTE)*, 9(1), 1–4. <https://doi.org/10.35940/ijrte.f7639.059120>

Hagiwara Y., & Tanaka T. (2020). *YuruGAN: Yuru-Chara Mascot Generator Using Generative Adversarial Networks With Clustering Small Dataset*.

<https://www.researchgate.net/publication/340776659>

Jin Y., Zhang J., Li M., Tian Y., Zhu H., & Fang Z. (2017). *Towards the Automatic Anime Characters Creation with Generative Adversarial Networks*. *Computer Vision and Pattern Recognition*. <https://www.researchgate.net/publication/319187018>

Karras, T., Laine, S., & Aila, T. (2021). *A style-based generator architecture for generative Adversarial Networks*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4217–4228. <https://doi.org/10.1109/tpami.2020.2970919>

Krohn, J., Beyleveld, G., & Bassens, A. (2020). *Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence*. Boston: Addison-Wesley.

Li, B., Zhu, Y., Wang, Y., Lin, C.-W., Ghanem, B., & Shen, L. (2021). Anigan: Style-

guided generative adversarial networks for unsupervised anime face generation. *IEEE Transactions on Multimedia*, 1–1. <https://doi.org/10.1109/tmm.2021.3113786>

Ojha, U., Li, Y., Lu, J., Efros, A., Lee, J., Shechtman, E., & Zhang, R. (2021). Few-shot

Image Generation via Cross-domain Correspondence. arXiv:2104.06820v1

Park, Bong-Won & Lee, Kun Chang. (2011). Exploring the value of purchasing online

game items. *Computers in Human Behavior*, 27(6).

<https://doi.org/10.1016/j.chb.2011.06.013>.

Tseng H., Jiang L., Liu C., Yang M., & Yang W. (2021) Regularizing Generative Adversarial

Networks under Limited Data. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7917-7927. doi: 10.1109/CVPR46437.2021.00783

Vavilala, V., & Forsyth, D. (2022). Controlled gan-based creature synthesis via a

challenging game art dataset - addressing the noise-latent trade-off. 2022

IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).

<https://doi.org/10.1109/wacv51458.2022.00019>