JASET KMUTNB Research Article

Journal of Applied Science and Emerging Technology (JASET)
Published by Faculty of Applied Science, KMUTNB, Thailand
doi: 10.14416/JASET.KMUTNB.2023.02.009
Vol. 22, No. 2 [2023]: e249246

Generative Adversarial Network สำหรับการสร้างตัวละคร DOTA2

Generative Adversarial Network for DOTA2 Character Generation

สรัลชัย อังควินิจวงศ์ พงศรัชฎ์ ชูทัย กฤตติพัฒน์ ชื่นพิทยาวุฒ ธีธัช นาคเสนีย์ และ ฐิติรัตน์ ศิริบวรรัตนกุล*

Saranchai Angkawinijwong, Pongsarat Chootai, Krittipat Chuenphitthayavut, Theethut Narksenee and Thitirat Siriborvornratanakul*

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ (นิด้า) กรุงเทพมหานคร 10240 ประเทศไทย

Graduate School of Applied Statistics, National Institute of Development Administration, Nida, Bangkok, Thailand

บทคัดย่อ

เกม DOTA2 เป็นเกมในอุตสาหกรรมเกมอีสปอร์ต (E-Sport Game) ที่ได้รับความนิยมและมีมูลค่าทางการตลาดสูง โดยเฉพาะหากสามารถพัฒนาลักษณะตัวละครในเกมให้มีความหลากหลายก็จะยิ่งเพิ่มรายได้ให้กับบริษัทและความพึงพอใจของผู้เล่นได้ จากการศึกษางานวิจัยในอดีตพบว่าการประยุกต์ใช้เครือข่ายขัดแย้งกำเนิด (Generative Adversarial Networks) หรือที่เรียกชื่อย่อว่า GAN มาใช้ในการสร้างภาพการ์ตูนหรือสร้างภาพบุคคลขึ้นมาใหม่นั้นได้ประสิทธิผลที่ดี งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาประสิทธิผลของการประยุกต์ใช้ StyleGAN ในการสร้างและออกแบบตัวละคร DOTA2 ขึ้นมาใหม่ในลักษณะการใช้ภาพจำนวนน้อยในการฝึกสอนต่อยอดผ่านการโต้ตอบข้ามโดเมน (Few-shot Image Generation via Cross-domain Correspondence)

ในงานวิจัยนี้ข้อมูลที่ใช้ในการฝึกสอน คือ รูปขนาดย่อของตัวละครฮีโร่เกมส์ DOTA2 ขนาด 256x144 พิกเซล จำนวน 111 รูป โดยภาพจะถูกสร้างขึ้นมาโมเดลทดลอง 3 ตัว โดยผลการสังเกตุโดยผู้วิจัยพบว่าโมเดลที่ประยุกต์มาจากโมเดล FFHQ สามารถดัดแปลงส่วนตาและจมูกของคนมาเป็นตาและจมูกของฮีโร่ได้อย่างถูกต้อง แต่ส่วนปากจะถูกลบหายไป ในขณะที่โมเดลอื่น ๆ สามารถสร้างภาพได้เพียงครึ่งส่วนแต่ยังมีความสอดคล้องกับลักษณะตัวละครอยู่ ทั้งนี้รูปภาพที่สร้างออกมาจากโมเดลทั้ง 3 มีสีที่ค่อนข้างเข้มและสดเหมือนกับรูปภาพของฮีโร่เกมส์ DOTA2 ในด้านผลลัพธ์เชิงคุณภาพผลการทดลองพบว่ากลุ่มตัวอย่างชื่นชอบ โมเดล FFHQ มากที่สุด ตามด้วยโมเดล Church และ โมเดล Horses ตามลำดับ โดยปัจจัยหลักที่กลุ่มตัวอย่างใช้ในการตัดสินใจเลือกโมเดลที่ชื่นชอบนั้นมาจากปัจจัยเรื่องกลิ่นไอของความเป็น DOTA2 ในภาพเป็นสำคัญ

คำสำคัญ: Generative Adversarial Networks, 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. Developing a variety of in-game character traits 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 StyleGAN application to be used in recreating and redesigning DOTA2 characters. To create new characters using a small number of DOTA2 original character images, we use a few-shot Image Generation via Cross-domain Correspondence. The dataset used in training includes 111 thumbnail images of DOTA2 hero characters. Three models were trained and experimented with. Manual observation showed that models adapted from the FFHQ model correctly adapted a person's eyes and nose to the character's eyes and nose; however, the mouth part was erased. The 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. As for the qualitative evaluation, it was found that users preferred the FFHQ model the most, followed by the Church model and the Horses model, respectively. The main consideration factors were how much the generated images are similar in style to DOTA2.

KEYWORDS: Generative Adversarial Networks, Deep Learning, DOTA2 , StyleGAN, Few-shot Image Generation

*Corresponding Author: thitirat@as.nida.ac.th

Received: 29/07/2022; Revised: 27/04/2023; Accepted: 14/06/2023

1. บทนำ

ในปัจจุบันอุตสาหกรรมเกมและเกมอีสปอร์ตกำลังเติบโตขึ้นอย่างรวดเร็วทั่วทุกมุมโลก อันเนื่องมาจากจุดเด่นหลายประการ เช่น ความสนุกสนาน ความตื่นเต้นท้าทาย นอกจากนี้การมีกฎและกติกาที่ชัดเจน ยังทำให้เกมอีสปอร์ตเปรียบเสมือนกีฬาชนิดหนึ่งที่สามารถจัดการแข่งขันขึ้นที่ใดก็ได้บนโลก โดยในการทำรายงานวิเคราะห์ตลาดของ Grand View Research (2020) พบว่า ในปี พ.ศ. 2563 อุตสาหกรรมเกมและเกมอีสปอร์ตมีส่วนแบ่งการตลาดอยู่ที่ 1.48 พันล้านเหรียญสหรัฐ อีกทั้งมีแนวโน้มจะเพิ่มสูงขึ้นเรื่อย ๆ คาดการณ์ว่าส่วนแบ่งการตลาดจะอยู่ที่ 6.81 พันล้านเหรียญสหรัฐ ในปี พ.ศ. 2570

หนึ่งในเกมอีสปอร์ตที่ได้รับความนิยมสูง ได้แก่ เกม DOTA2 ซึ่งบริษัทวาล์วคอร์ปอเรชัน (Valve) ได้พัฒนาขึ้น โดยมีผู้เล่นจำนวนมากกว่า 10 ล้านคนทั่วโลก (Clement, 2021) อีกทั้งมีการจัดการแข่งขันในระดับภูมิภาคและระดับโลกอย่างต่อเนื่อง ในเกม DOTA2 นี้ผู้เล่นสามารถเลือกใช้งานตัวละครได้มากถึง 121 ตัวละคร โดยตัวละครจะถูกจำแนกออกเป็น 3 กลุ่ม ได้แก่ (1) ตัวละคร Strength ซึ่งมีภาพลักษณ์แข็งแรง ดุดัน น่าเกรงขาม มีพลังทางกายภาพสูง (2) ตัวละคร Agility ซึ่งมีภาพลักษณ์คล่องแคล่ว ว่องไว ปราดเปรียว และ (3) ตัวละคร Intelligence ซึ่งมีภาพลักษณ์ที่บอบบาง อ่อนแอ ทรงความรู้ เป็นผู้ใช้เวทมนตร์ เป็นต้น

จากการทบทวนวรรณกรรม ผู้วิจัยพบว่าการสร้างตัวละครใหม่เพิ่มเติมขึ้นมาในเกมนั้น นอกจากจะส่งผลให้ผู้เล่นมีทางเลือกในการเล่นมากขึ้นแล้ว หากตัวละครใหม่นั้นถูกออกแบบมาอย่างสวยงาม มีกลิ่นไอความเป็นตัวละครของ DOTA2 ตามประเภทของตัวละครที่กล่าวมาข้างต้นอย่างถูกต้อง ก็จะยิ่งส่งผลให้ผู้เล่นตัดสินใจจ่ายเงินซื้อตัวละครนั้นมาใช้งานกันมากขึ้น เนื่องจากเอกลักษณ์ของตัวละครมีความสัมพันธ์เชิงบวกกับความตั้งใจที่จะซื้ออุปกรณ์และสิ่งต่าง ๆในเกม (Park & Lee, 2011) ไม่เพียงเท่านี้ความคาดหวังที่ตัวละครจะสามารถเติบโตเปลี่ยนแปลงไปในทางที่ดีขึ้นยังมีผลทางบวกต่อความภักดีของผู้เล่นเกมออนไลน์นั้นอีกด้วย (Alghifari & Halim, 2020) อย่างไรก็ตาม การสร้างตัวละครใหม่ขึ้นมานั้นมีค่าใช้จ่ายในการออกแบบที่ค่อนข้างสูง อีกทั้งมีความเสี่ยงว่าตัวละครใหม่ที่ลงทุนออกแบบมาอาจไม่เป็นที่ถูกใจผู้เล่นโดยรวม

ด้วยความก้าวหน้าของเทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) ในปัจจุบัน ทำให้ผู้ออกแบบและพัฒนาตัวละครสามารถนำศิลปินปัญญาประดิษฐ์ (Artificial Intelligence Artist หรือ AI artist) มาใช้ช่วยในการออกแบบตัวละครใหม่อย่างรวดเร็วได้ โดยศิลปินปัญญาประดิษฐ์ถือเป็นเจนเนอเรทีฟโมเดล (Generative Model) รูปแบบหนึ่งซึ่งมีสถาปัตยกรรมให้เลือกใช้ได้หลายแบบในปัจจุบัน สำหรับงานวิจัยชิ้นนี้ผู้วิจัยเลือกใช้สถาปัตยกรรมของเครือข่ายขัดแย้งกำเนิด (Generative Adversarial Network หรือ GAN) (Goodfellow et al., 2014) ซึ่งภายในประกอบด้วยโครงข่ายย่อยอีก 2 ส่วน ได้แก่ ส่วนของเจนเนอเรเตอร์ (Generator) และส่วนของดิสคริมิเนเตอร์ (Discriminator) โดยเครือข่ายขัดแย้งกำเนิดนี้มีความสามารถในการสร้างภาพเหมือนและภาพการ์ตูนได้ดี (Krohn et al., 2020) อีกทั้งยังได้รับความนิยมอย่างสูงในงานการสร้างรูปภาพเสมือนจริง (Photorealistic image) ที่มีความละเอียดและความสมจริงสูงด้วย

ในบรรดาเครือข่ายขัดแย้งกำเนิดที่ถูกนำเสนอขึ้นมามากมายนั้น StyleGAN (Karras et al., 2021) ของบริษัท NVIDIA กล่าวได้ว่าเป็นเครือข่ายขัดแย้งกำเนิดที่โดดเด่นมากที่สุด เนื่องจากสามารถสกัดคุณสมบัติระดับสูง เช่น ท่าโพส ท่าทาง สเกล ส่วนประกอบในภาพ ได้อย่างอัตโนมัติ อีกทั้งมีการปรับแต่งในส่วนของเจนเนอเรเตอร์ทำให้สามารถสร้างภาพผลลัพธ์ที่คมชัดและมีสัดส่วนของของภาพที่ถูกต้องกว่าการใช้เจนเนอเรเตอร์แบบดั้งเดิม นอกจากนี้ StyleGAN ยังสามารถปรับสไตล์ (Latent) เพื่อผสมและจับคู่สไตล์ของภาพสองภาพเข้าด้วยกันได้ด้วย เหล่านี้ทำให้ StyleGAN เป็นหนึ่งในเครือข่ายขัดแย้งกำเนิดยอดนิยมที่นักพัฒนานิยมนำไปต่อยอดเพื่อสร้างรูปภาพประเภทต่าง ๆ ด้วยเหตุนี้ คณะผู้วิจัยจึงเล็งเห็นถึงความเป็นไปได้ของการนำ StyleGAN มาทำการทดลองเรียนรู้ และสร้างตัวละคร DOTA2 ใหม่โดยให้อ้างอิงกับเอกลักษณ์และสไตล์ของตัวละครที่มีอยู่ปัจจุบันในเกม DOTA2 ซึ่งหากทำได้สำเร็จก็น่าจะช่วยลดต้นทุนการสร้างตัวละครใหม่และส่งเสริมยอดขายโดยรวมของเกม DOTA2 ได้เป็นอย่างดี

2. ทบทวนวรรณกรรม

ในหัวข้อนี้ผู้วิจัยทำการทบทวนวรรณกรรมที่เกี่ยวข้องกับการนำเครือข่ายขัดแย้งกำเนิดไปใช้ในการสร้างภาพสองมิติ เพื่อพิจารณาถึงความเป็นไปได้ จุดเด่น และจุดด้อย

ของเครือข่ายขัดแย้งกำเนิด ก่อนจะสรุปประเด็นที่น่าสนใจเกี่ยวกับเครือข่ายขัดแย้งกำเนิดในบริบทของการนำไปสร้างภาพตัวละครใหม่ของเกม DOTA2 ในงานวิจัยนี้ต่อไป

หากพิจารณาเปรียบเทียบเอกลักษณ์เชิงภาพระหว่างภาพถ่ายและภาพการ์ตูนนั้น ภาพการ์ตูนจะมีเอกลักษณ์ในแง่ของลักษณะลายเส้นเสียมาก ในขณะที่จะมีความสมจริงและรายละเอียดของพื้นผิว (Texture) ในองค์ประกอบต่าง ๆ ของภาพน้อยกว่าภายถ่าย ที่ผ่านมามีการศึกษามากมายที่ทดลองนำเครือข่ายขัดแย้งกำเนิดมาใช้ในการสร้างภาพการ์ตูน เช่น Li et al. (2021) นำเสนอเฟรมเวิร์กใหม่ที่เรียกว่า AniGAN ซึ่งสามารถสังเคราะห์หน้าตาตัวละครอนิเมะคุณภาพสูงได้ ในรายละเอียดมีการใช้เจนเนอเรเตอร์ที่สามารถถ่ายโอนรูปแบบ สี พื้นผิว ให้กลายเป็นหน้าอนิเมะได้พร้อม ๆ กัน โดยสถาปัตยกรรมที่ใช้มีการใช้โครงข่ายส่วนการเข้ารหัส (Encoder) แยกระหว่างรูปภาพพื้นผิวหน้าคนกับภาพสไตล์อนิเมะ และใช้โครงข่ายส่วนการถอดรหัส (Decoder) แปลงภาพให้กลายเป็นรูปภาพสไตล์อนิเมะที่สะท้อนจากรูปภาพหน้าคน เนื่องด้วยใบหน้าของอนิเมะและหน้าของคนมีองค์ประกอบของรูปภาพที่คล้ายคลึงกัน ดังนั้นโครงข่ายดิสคริมิเนเตอร์ในที่นี้จึงมีหน้าที่ 2 อย่าง ได้แก่ (1) การแยกแยะภาพจริงและภาพไม่จริงของใบหน้าอนิเมะ และ (2) การแยกแยะภาพจริงและไม่จริงของใบหน้ามนุษย์ ชุดข้อมูลที่ใช้ในงานวิจัยนี้ได้แก่ Selfie2anime และ Face2anime จากผลการทดลองสามารถสรุปได้ว่าวิธีที่นำเสนอได้ผลลัพธ์ที่ดีกว่าเมื่อเทียบกับวิธีการก่อนหน้า Back (2021) เสนอวิธีการชื่อ Cartoon-StyleGAN ซึ่งใช้เทคนิคการฝึกสอนแบบไม่มีผู้สอน (Unsupervised learning) ในงานการแปรภาพ (Image-to-Image translation หรือ I2I) โดยการทำ I2I แบบไม่มีผู้สอนนั้นใช้การถ่ายโอนการเรียนรู้จากโมเดล StyleGAN2 ที่ถูกฝึกสอนไว้ล่วงหน้า (Pre-trained StyleGAN2) และมีการใช้วิธีการใหม่เพื่อรักษาโครงสร้างของภาพต้นฉบับและสร้างภาพที่เหมือนจริงในโดเมนเป้าหมายด้วย ผลการทดลองแสดงให้เห็นว่าวิธีการเหล่านี้มีประสิทธิภาพในการทำให้รูปภาพต้นฉบับและรูปภาพเป้าหมายมีความคล้ายคลึงกัน และช่วยให้สร้างภาพได้สมจริงยิ่งขึ้น

Vavilala & Forsyth (2022) อภิปรายว่าเนื่องด้วย StyleGAN ถูกใช้อย่างหลากหลายในงานปรับแต่งรูปภาพต่าง ๆ รวมทั้งในการปรับแต่งการ์ดเกมด้วย ทว่าการปรับแต่งนี้ยังไม่สามารถทำให้เกิดความแตกต่างจากภาพต้นฉบับอย่างมีนัยสำคัญได้ ดังนั้นงานนี้จึงเสนอการปรับแต่ง StyleGAN โดยใช้ชุดข้อมูลจาก The Yu-Gi-Oh Card Art Dataset ซึ่งมีข้อมูลอยู่ 11,000 ภาพ เริ่มจากการทำความสะอาดข้อมูลโดยใช้สายตาคัดแยกและคัดออกประมาณ 500 ภาพเนื่องจากไม่สามารถระบุการ์ดนั้นได้ จากนั้นทำการย่อขนาดภาพให้เป็นขนาด 256 พิกเซล และ 512 พิกเซล โดยงานนี้ได้ทำการฝึกสอนด้วยภาพขนาด 512 พิกเซลสำหรับการ์ดประเภทมอนสเตอร์โดยใช้เวลา 227 ชั่วโมงในการฝึกสอนสำหรับรูปภาพ 25 ล้านภาพ ผลลัพธ์ที่ดีที่สุดได้ค่าตัวชี้วัด FID ที่ 10.73 และจากการทดลองสามารถสรุปผลได้ว่า Style-GAN2 สามารถสร้างการ์ดที่น่าดึงดูดและมีความแตกต่างจากภาพต้นฉบับที่ใช้ฝึกสอนอย่างมีนัยสำคัญได้

จากตัวอย่างงานวิจัยที่กล่าวไปข้างต้นจะเห็นว่าเครือข่ายขัดแย้งกำเนิดสามารถถูกนำไปฝึกสอนให้สร้างรูปภาพลายเส้นแบบของการ์ตูนหรืออนิเมะได้ โดยเฉพาะเครือข่ายขัดแย้งกำเนิดในกลุ่มของ StyleGAN ซึ่งเป็นที่นิยมในหมู่นักวิจัยสำหรับการนำมาทดลองฝึกสอนต่อยอด อย่างไรก็ตาม ปัญหาของการฝึกสอนเครือข่ายขัดแย้งกำเนิดขนาดใหญ่เหล่านี้คือจำนวนของข้อมูลฝึกสอน โดยเฉพาะหากเป็นการฝึกสอนตั้งแต่ศูนย์ (Train from scratch) จะยิ่งต้องการรูปภาพสำหรับฝึกสอนจำนวนมาก อีกทั้งต้องใช้ทรัพยากรคอมพิวเตอร์สูงและใช้เวลาในการฝึกสอนนานตามไปด้วย จึงอาจไม่เหมาะสมสำหรับเป้าหมายของงานวิจัยนี้ที่ต้องการทำให้กระบวนการออกแบบตัวละครใหม่มีความรวดเร็วขึ้นกว่าปัจจุบัน ตัวอย่างปริมาณข้อมูลฝึกสอนที่ใช้ในการฝึกสอนเครือข่ายขัดแย้งกำเนิดสำหรับสร้างภาพการ์ตูน เช่น งานของ Guruprasad et al. (2020) ทำการฝึกสอนโมเดลเพื่อสร้างใบหน้าคนในรูปแบบภาพการ์ตูนโดยใช้เครือข่ายขัดแย้งกำเนิดแบบพื้นฐานที่ฝึกสอนบนชุดข้อมูล CartoonSet100K ซึ่งเป็นชุดข้อมูลที่ประกอบด้วยใบหน้าคนในรูปแบบภาพการ์ตูนที่มีขนาดรูปภาพ 500x500 พิกเซล จำนวนหนึ่งแสนภาพ หรืองานของ Jin et al. (2017) ที่ใช้ DRAGAN (Deep Regret Analytic Generative Adversarial Networks) ร่วมกับการประยุกต์ใช้โครงสร้างโมเดล SRResNet เป็นส่วนประกอบในเจนเนอเรเตอร์และดิสคริมิเนเตอร์ในการสร้างเครือข่ายขัดแย้งกำเนิดสำหรับเรียนรู้ที่จะสร้างหน้าตัวละครอนิเมะขึ้นมาใหม่ โดยข้อมูลรูปภาพส่วนใบหน้าตัวละครจากเกมตั้งแต่ปี 2005 ที่ได้จากเว็บไซต์จำนวน 31,255 รูปภาพ ผลลัพธ์ของการสร้างภาพพบว่าโมเดลสามารถสร้างรูปภาพได้ดีในอนิเมะที่มีผมสีบลอนด์และตาสีฟ้า แต่จะทำได้ไม่ดีนักในอนิเมะ

ที่สวมแว่นหรือมีการม้วนผม เนื่องจากเป็นคุณลักษณะที่พบได้น้อยในชุดข้อมูลที่ใช้ฝึกสอน

ในส่วนของเทคนิคที่ถูกนำเสนอเพื่อแก้ปัญหาปริมาณข้อมูลฝึกสอนที่มีน้อยเกินไปสำหรับจะใช้ฝึกสอนเครือข่ายขัดแย้งกำเนิดนั้นก็มีหลากหลาย เช่น Hagiwara & Tanaka (2020) ทดลองใช้เงื่อนไขในระดับคลาส (Class condition) กับชุดข้อมูลที่นำไปฝึกสอนซึ่งมีจำนวนน้อยไม่เพียงพอต่อการฝึกสอนด้วยเทคนิคการฝึกสอนตามปกติ ดังนั้นจึงมีการใช้เทคนิคการแบ่งกลุ่มข้อมูล (Clustering) และการเพิ่มปริมาณข้อมูลรูปภาพฝึกสอน (Data Augmentation) มาช่วยทำให้สามารถเพิ่มปริมาณข้อมูลฝึกสอนขึ้นถึง 5 เท่า จากเดิมที่มีภาพขนาด 144 x 128 พิกเซลจำนวน 4,018 ภาพก็กลายเป็น 20,090 ภาพได้ จากนั้นมีการทดลองนำภาพเหล่านี้ไปผ่านการกระบวนการสกัดฟีเจอร์ (Feature Extraction) ด้วยวิธีการต่าง ๆ แล้วนำฟีเจอร์ทั้งหมดที่คำนวณได้ไปใช้ทดลองสร้างเจนเนอเรทีฟโมเดลอีกทีหนึ่ง

Tseng et al. (2021) นำเสนอวิธีการฝึกสอนมาตรฐานสำหรับเครือข่ายขัดแย้งกำเนิดภายใต้ข้อมูลที่จำกัด ในการทดลองมีการใช้ชุดข้อมูล CIFAR 10/100 และ ImageNet และทำการทดลองสร้างภาพต่าง ๆ ด้วยเครือข่ายขัดแย้งกำเนิดต่างชนิดกันเพื่อสาธิตประสิทธิภาพของเทคนิคการฝึกสอนที่นำเสนอไปดังกล่าวในประเด็นที่ว่า (1) สามารถปรับปรุงประสิทธิภาพของเครือข่ายขัดแย้งกำเนิดภายใต้ข้อจำกัดการตั้งค่าข้อมูล และ (2) สามารถนำไปใช้กับวิธีการเสริมข้อมูลหรือวิธีการเพิ่มปริมาณข้อมูลภาพฝึกสอนเพื่อเพิ่มประสิทธิภาพการทำงานต่อไป

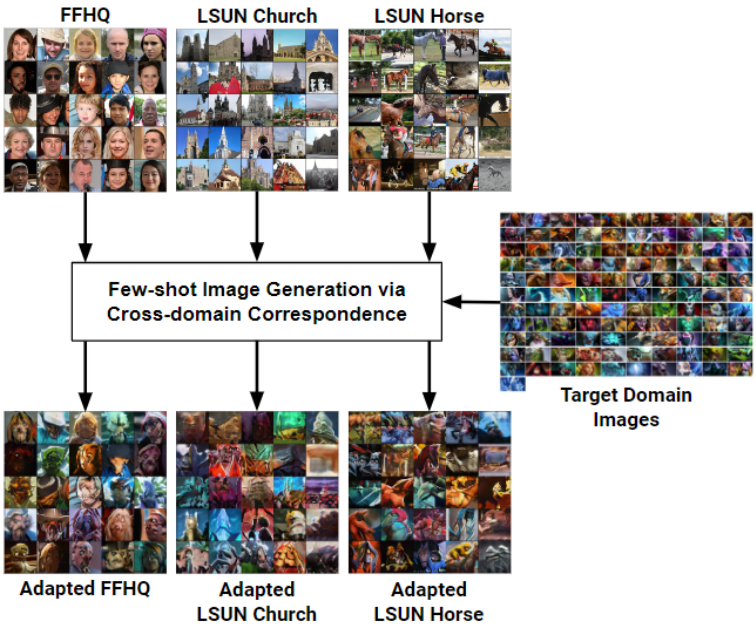
Ojha et al. (2021) ได้ศึกษาการสร้างเครือข่ายขัดแย้งกำเนิดในลักษณะการสร้างภาพช็อตจำนวนน้อยผ่านการโต้ตอบข้ามโดเมน (Few-shot Image Generation via Cross-domain Correspondence) งานนี้กล่าวถึงปัญหาของการฝึกสอนเครือข่ายขัดแย้งกำเนิดในสถานการณ์ที่โดเมนเป้าหมายมีตัวอย่างข้อมูลสำหรับฝึกสอนจำกัด ทำให้เกิดปัญหาโอเวอร์ฟิต (Overfit) ตามมาได้ง่าย เพื่อแก้ปัญหาดังกล่าวงานนี้เลือกใช้ประโยชน์จากโดเมนต้นทางขนาดใหญ่สำหรับการเทรนล่วงหน้า (Pretraining) บนข้อมูลที่หลากหลายและมีจำนวนมาก จากนั้นจึงทำการถ่ายโอนความหลากหลายของข้อมูลดังกล่าวจากต้นทางมาสู่เป้าหมายในลักษณะที่สามารถรักษาความเหมือนและความแตกต่างที่สัมพันธ์กันระหว่างข้อมูลโดยใช้ระยะห่างของการสูญเสียคงที่ในโดเมนข้ามแบบใหม่ (Novel Cross-Domain Distance Consistency Loss) จากผลการทดลองทั้งเชิงคุณภาพและเชิงปริมาณ ผู้วิจัยได้สรุปว่าการสร้างภาพโมเดลช็อตจำนวนน้อยจะค้นหาการโต้ตอบโดยอัตโนมัติระหว่างโดเมนต้นทางและปลายทาง และสร้างภาพที่มีความหลากหลายและสมจริงได้มากกว่าวิธีการในอดีต

จากการทบทวนวรรณกรรมที่กล่าวมาข้างต้นทั้งหมดสามารถแสดงให้เห็นถึงประสิทธิภาพ ประสิทธิผล และปัญหาด้านต่าง ๆ ของการประยุกต์ใช้เครือข่ายขัดแย้งกำเนิดได้เป็นอย่างดี นอกจากนี้ยังเป็นที่สังเกตได้ว่านักวิจัยนิยมนำโมเดล StyleGAN มาใช้ในการสร้างภาพการ์ตูน ภาพมาสคอต หรือสร้างภาพบุคคลขึ้นใหม่อย่างแพร่หลาย คณะผู้วิจัยจึงเล็งเห็นความเป็นไปได้ในการประยุกต์ใช้ StyleGAN ในการสร้างและออกแบบตัวละคร DOTA2 ขึ้นมาใหม่ โดยจะนำโมเดล StyleGAN ที่ถูกฝึกสอนล่วงหน้าไว้อย่างดีแล้วมาปรับปรุง และฝึกสอนเพิ่มเติมโดยใช้ภาพตัวละคร DOTA2 เพื่อปรับให้แบบจำลอง StyleGAN สามารถสร้างภาพตัวละครที่มีเอกลักษณ์ตรงตามแบบของเกม DOTA2 ได้ต่อไป ในงานวิจัยนี้เนื่องจากจำนวนภาพฝึกสอน DOTA2 ที่คณะผู้วิจัยสามารถรวบรวมมาได้มีจำนวนน้อย เพื่อป้องกันปัญหาโอเวอร์ฟิตที่อาจตามมาได้ภายหลัง ในงานวิจัยนี้จึงจะนำแนวทางแก้ปัญหาโอเวอร์ฟิตของ Ojha et al. (2021) สำหรับการสร้างภาพช็อตจำนวนน้อยผ่านการโต้ตอบข้ามโดเมน (Few-shot Image Generation via Cross-domain Correspondence) มาประยุกต์ใช้ด้วย

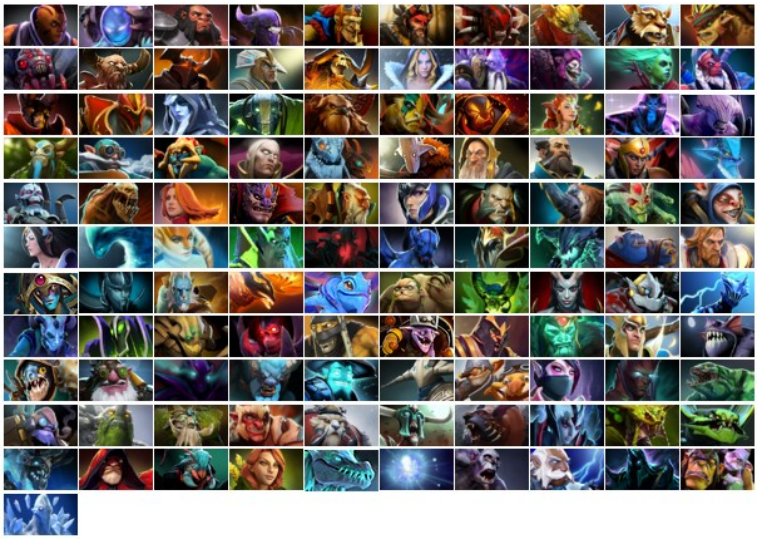
3. วิธีการทดลอง

เทคนิคการสร้างภาพจากข้อมูลฝึกสอนจำนวนน้อย (Few-shot Image Generation)

งานวิจัยนี้ประยุกต์ใช้เทคนิคการเรียนรู้จากงานวิจัย (Ojha et al., 2021) ซึ่งเป็นเทคนิคที่ทำให้สามารถนำเจนเนอเรเตอร์ที่ถูกฝึกสอนไว้ล่วงหน้า (Pre-trained Generator) มาประยุกต์เพื่อสร้างภาพออกมาในรูปแบบที่ผู้ใช้งานต้องการได้ โดยจุดเด่นของวิธีนี้คือสามารถใช้รูปภาพฝึกสอนเพียงแค่ 1 ภาพก็สามารถสร้างผลลัพธ์ออกมาได้เป็นที่น่าพอใจ อย่างไรก็ตามจำนวนรูปภาพฝึกสอนที่มากขึ้นก็จะช่วยให้ผลลัพธ์ดีขึ้นและมีความหลากหลายมากขึ้นด้วย เทคนิคที่งานวิจัยของ (Ojha et al., 2021) ใช้คือเทคนิค Cross-domain Correspondence ที่ทำให้โมเดลสามารถส่งผ่านความแตกต่างที่เกิดขึ้นในฟีเจอร์ของโดเมนต้นทาง (Source Domain) ไปยังโดเมนปลายทาง (Domain Destination) ได้ นอกจากนี้เทคนิคนี้ยังพยายามทำให้การกระจายตัวของฟีเจอร์ของโดเมน



รูปที่ 1 ภาพรวมของการทดลองในงานวิจัยนี้



รูปที่ 2 รูปฮีโร่เกมส์ DOTA2 ที่ใช้ในการเรียนรู้ จำนวน 111 รูป (ที่มาของภาพจาก (AucT, 2013))

ตารางที่ 1 พารามิเตอร์ที่ใช้ในการฝึกสอนต่อยอดของโมเดล StyleGANv2 ที่ถูกฝึกสอนไว้ล่วงหน้า

Parameter	Value
Iteration	25000
Batch Size	4
Generated Image Size	256x256 pixel
Latent Dimension	512
Learning Rate	0.002
Augmentation	**True**

เป้าหมายมีลักษณะใกล้เคียงกับการกระจายตัวของฟีเจอร์ของโดเมนต้นทางอีกด้วย

เทคนิคของ (Ojha et al., 2021) ต้องการอินพุต 2 อย่าง ได้แก่ (1) โมเดลที่ถูกฝึกสอนไว้ล่วงหน้า (Pre-trained Model) ซึ่งผู้วิจัยเลือกใช้เป็นโมเดล StyleGANv2 (Karras et al., 2020) และ (2) ภาพจากโดเมนเป้าหมายอย่างน้อย 1 ภาพ ซึ่งภาพโดเมนเป้าหมายจะเป็นภาพที่ต้องการให้โมเดลที่ถูกฝึกสอนไว้ล่วงหน้าทำการเรียนรู้และประยุกต์โมเดลไปสร้างภาพในรูปแบบของโดเมนเป้าหมายที่กำหนด หลังจากการฝึกสอนผลลัพธ์ที่จะได้คือโครงข่ายส่วนของเจนเนอเรเตอร์ที่สามารถสร้างรูปภาพในลักษณะเดียวกันกับโมเดลที่ถูกฝึกสอนไว้ล่วงหน้า แต่มีการประยุกต์ในโดเมนเป้าหมายที่ต้องการให้เรียนรู้เพิ่มเข้าไปด้วย เช่น กำหนดโมเดลที่ถูกฝึกสอน

ตารางที่ 2 ผลลัพธ์ที่ได้จาก Adapted Generator ของโมเดล FFHQ, LSUN Church และ LSUN Horse ตั้งแต่ Iteration ที่ 0 ถึง 10,000

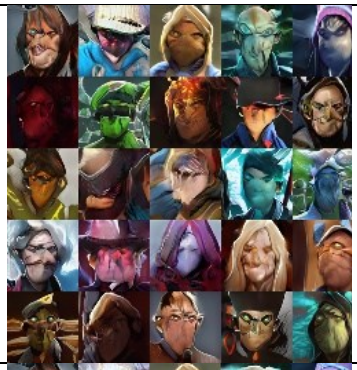
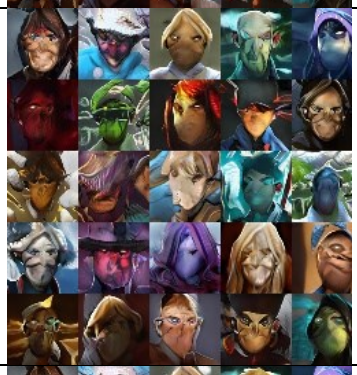
Iteration	Model FFHQ	Model LSUN Church	Model LSUN Horse
0			
5000			
10000			

ไว้ล่วงหน้าเป็นโมเดลที่ถูกฝึกสอนไว้สำหรับสร้างภาพใบหน้าคน และกำหนดภาพโดเมนเป้าหมายคือรูปวาดใบหน้าคนในสไตล์ของศิลปิน A ผลลัพธ์ที่ได้จากการฝึกสอนโมเดล คือ โมเดลที่สามารถสร้างรูปวาดใบหน้าคนในรูปแบบและสไตล์การวาดของศิลปิน A ได้ โดยยังคงรักษาโครงสร้างใบหน้าที่สำคัญจากโมเดลที่ถูกฝึกสอนไว้ล่วงหน้าได้และไม่เกิดการโอเวอร์ฟิตกับภาพโดเมนเป้าหมายที่กำหนดให้เรียนรู้ แม้ว่าภาพที่ใช้ฝึกสอนจะมีรูปภาพเพียง 1 ภาพก็ตาม

เนื่องจากข้อจำกัดทางด้านจำนวนรูปภาพของฮีโร่ในเกมส์ DOTA2 ที่มีจำนวนไม่มาก ในการศึกษานี้ คณะผู้วิจัยจึงนำเทคนิค Few-shot Image Generation via Cross-domain Correspondence (Ojha et al., 2021) มาใช้กับโมเดลที่ถูกฝึกสอนไว้ล่วงหน้าด้วย โดยงานวิจัยนี้จะทำการฝึกสอนและเปรียบเทียบผลลัพธ์การสร้างภาพฮีโร่ของเกม DOTA2 จากผลลัพธ์ของโมเดลที่ถูกฝึกสอนไว้ล่วงหน้าเริ่มต้นที่แตกต่างกัน 3 โมเดลซึ่งทั้งหมดเป็นโมเดลสถาปัตยกรรมของ StyleGANv2 ได้แก่ (1) โมเดล FFHQ (Flickr-Faces-HQ) ที่ใช้สำหรับสำหรับสร้างภาพใบหน้าคน (2) โมเดล LSUN (Large-scale Scene Understanding) Church ที่ใช้สำหรับสร้างภาพโบสถ์ และ (3) โมเดล LSUN Horse ที่ใช้สำหรับสำหรับสร้างภาพม้า โดยผู้วิจัยกำหนดให้โมเดลที่ถูกฝึกสอนล่วงหน้ามาบนชุดข้อมูลที่แตกต่างกันเหล่านี้ ถูกนำมาฝึกสอนต่อยอดโดยมีภาพโดเมนเป้าหมายร่วมกันคือรูปภาพฮีโร่ DOTA2 จำนวน 111 ภาพ ทั้งนี้โมเดลทุกประเภทจะถูกกำหนดเงื่อนไขพารามิเตอร์ในการเรียนรู้ และรูปภาพโดเมนเป้าหมายที่เหมือนกันทุกประการ แตกต่างกันเพียงโมเดลที่ถูกฝึกสอนไว้ล่วงหน้าที่ไม่เหมือนกันเท่านั้น

รูปภาพที่ 1 สรุปภาพรวมของการทดลองในงานวิจัยนี้ เมื่อโมเดล FFHQ โมเดล LSUN Church และโมเดล LSUN Horse หมายถึงโมเดล StyleGANv2 ตั้งต้นที่ถูกฝึกสอนไว้ล่วงหน้าบนภาพถ่ายที่ไม่ใช่ภาพการ์ตูนและไม่เกี่ยวข้องกับตัวละคร DOTA2 แต่อย่างใด ซึ่งโมเดลตั้งต้น

ตารางที่ 3 ผลลัพธ์ที่ได้จาก Adapted Generator ของโมเดล FFHQ, LSUN Church และ LSUN Horse ตั้งแต่ Iteration ที่ 15,000 ถึง 25,000 ต่อเนื่องจากตารางที่ 2

Iteration	Model FFHQ	Model LSUN Church	Model LSUN Horse
15000			
20000			
25000			

ทั้งสามนี้จะถูกงานวิจัยนี้นำมาฝึกสอนต่อยอดด้วยรูปภาพฮีโร่ DOTA2 จำนวน 111 รูป (Target Domain Images) ด้วยวิธี Few-shot Image Generation via Cross-domain Correspondence ด้วยการกำหนดค่าและเงื่อนไขพารามิเตอร์ในการเรียนรู้ดังแสดงในตารางที่ 1 ทำให้ได้ผลลัพธ์ของการฝึกสอนเป็นโมเดลใหม่ 3 ตัวชื่อโมเดล Adapted FFHQ โมเดล Adapted LSUN Church และโมเดล Adapted LSUN Horse ที่ต่างก็สามารถสร้างภาพตัวละครฮีโร่ในเกม DOTA2 ออกมาได้

ชุดข้อมูล (Dataset)

ในงานวิจัยนี้รูปภาพใบหน้าตัวละครฮีโร่ของเกมส์ DOTA2 ขนาด 256x144 พิกเซล จำนวน 111 รูป (AucT, 2013) ดังตัวอย่างแสดงในภาพที่ 2 จะถูกกำหนดให้เป็นภาพโดเมนเป้าหมายสำหรับให้โมเดลทั้ง 3 โมเดลที่ถูกฝึกสอนไว้ล่วงหน้าใช้ในการเรียนรู้ โดยรูปภาพจำนวน 111 รูปถือว่าเป็นจำนวนที่มากพอสำหรับการประยุกต์ใช้ในการฝึกสอนต่อยอดจากโมเดลที่ถูกฝึกสอนไว้ล่วงหน้าโดยใช้เทคนิค Cross-domain Correspondence อ้างอิงจากการทดลองในงานวิจัย (Ojha et al., 2021) ที่พบว่าจำนวนรูปภาพโดเมนเป้าหมายจำนวน 10 ภาพ ก็ให้ผลลัพธ์ที่น่าประทับใจแล้ว

การทดลองเชิงเปรียบเทียบ (Comparative Experiment)

เหตุผลที่ผู้วิจัยทำการเปรียบเทียบผลลัพธ์จากโมเดลเริ่มต้น 3 ตัวซึ่งถูกฝึกสอนมาให้สร้างภาพต่างชนิดกันอย่างชัดเจน ได้แก่ โมเดล FFHQ ที่สร้างภาพใบหน้าคน โมเดล LSUN Church ที่สร้างภาพโบสถ์ และโมเดล LSUN Horse ที่สร้างภาพม้านั้น เพราะต้องการเปรียบเทียบผลลัพธ์ที่ได้จากการประยุกต์โมเดลที่เป็นตัวแทนของส่วนประกอบหลักที่สำคัญในเกมที่ต่างกันว่าจะมีผลลัพธ์เป็นอย่างไร สามารถสร้างส่วนประกอบต่างประเภทในเกมได้ครบถ้วนหรือไม่ กล่าวคือ โมเดล FFHQ เป็นตัวแทนของคนหรือฮีโร่ในเกมส์ โมเดล LSUN Church เป็นตัวแทนของสิ่งก่อสร้างในเกมส์ เช่น บ้าน

ร้านค้า หรือป้อมปราการ และโมเดล LSUN Horse เป็นตัวแทนของสัตว์ในเกมส์ เช่น สัตว์ที่ใช้ส่งไอเทมหรือสัตว์ที่เป็นยานพาหนะของฮีโร่บางประเภท ทั้งนี้การทดลองในงานวิจัยนี้ถูกทำบนแพลตฟอร์ม Google Colab Pro ที่ใช้ฮาร์ดแวร์คือ CPU: Intel® Xeon® CPU@2.00GHz, GPU: Tesla P100, RAM: 13GB

4. ผลการทดลอง

ในการศึกษานี้จะแสดงผลลัพธ์ที่ได้จากการสร้างภาพโดยโครงข่ายส่วนของเจนเนอเรเตอร์ซึ่งอยู่ภายในโมเดลที่ผ่านการเรียนรู้ด้วยวิธี Few-shot Image Generation via Cross-domain Correspondence (Ojha et al., 2021) ตามโดเมนเป้าหมายซึ่งก็คือภาพฮีโร่ DOTA2 จำนวน 111 ภาพ โดยมีโมเดลต้นแบบคือ StyleGANv2 ที่ถูกฝึกสอนล่วงหน้าไว้บนชุดข้อมูลที่ต่างกัน 3 โมเดลตามที่ได้กล่าวไปในหัวข้อก่อนหน้า โดยจะแสดงผลลัพธ์ทุก ๆ 5,000 รอบ (Iteration) ของการฝึกสอนเริ่มตั้งแต่รอบที่ 0 ที่โมเดลยังไม่มีการเรียนรู้ต่อยอดจากเดิม จนกระทั่งถึงรอบที่ 25,000 โดยผลลัพธ์แสดงดังตารางที่ 2 และ 3 ต่อเนื่องกัน

การเปรียบเทียบเชิงปริมาณ (Quantitative comparison)

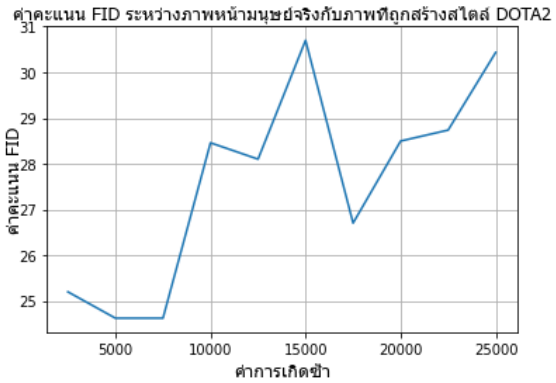
จากภาพผลการทดลองในตารางที่ 2 และ 3 พบว่าโมเดลที่ประยุกต์มาจากโมเดล FFHQ สามารถดัดแปลงส่วนตาและจมูกของคน มาเป็นตาและจมูกของฮีโร่ได้อย่างถูกต้อง แต่ส่วนปากจะถูกลบออกไป ในส่วนผลลัพธ์ที่ประยุกต์มาจากโมเดล LSUN Church แสดงให้เห็นว่าจำนวนภาพที่ถูกสร้างออกมาจะสามารถคงรูปร่างโบสถ์ไว้ได้เพียงครึ่งหนึ่ง และยังพบว่าโมเดลพยายามแทนที่หน้าต่างของโบสถ์ด้วยดวงตาของฮีโร่ และส่วนผลลัพธ์ที่ประยุกต์มาจากโมเดล LSUN Horse ก็มีผลลัพธ์ใกล้เคียงกันกับโมเดลจาก LSUN Church กล่าวคือภาพผลลัพธ์ที่ถูกสร้างออกมาจะสามารถคงรูปร่างม้าไว้ได้เพียงครึ่งหนึ่ง ทั้งนี้รูปภาพที่โมเดลทั้งสามสร้างออกมาจะมีสีที่ค่อนข้างเข้มและสด ซึ่งเป็นลักษณะเดียวกันกับรูปภาพของฮีโร่เกมส์ DOTA2 ที่กำหนดให้โมเดลทำการเรียนรู้นั่นเอง กล่าวได้ว่า ผู้วิจัยสามารถประยุกต์เทคนิคการเรียนรู้ตามแนวทางของ Ojha et al. (2021) ซึ่งเป็นเทคนิคที่ทำให้เจนเนอเรเตอร์ที่ถูกฝึกสอนไว้ล่วงหน้าไม่เกิดโอเวอร์ฟิตมาสร้างภาพตัวละคร DOTA2 ได้เป็นผลสำเร็จ

นอกจากนี้ผู้วิจัยยังทำการประเมินผลลัพธ์เชิงปริมาณด้วยเมตริก (Frechet Inception Distance (FID)(Heusel et al., 2017)) ซึ่งเป็นตัวเลขคะแนนที่นิยมใช้ในการประเมินคุณภาพของภาพสังเคราะห์ในงานสังเคราะห์ภาพ โดย FID จะคำนวณระยะห่างระหว่างฟีเจอร์เวกเตอร์ (Feature) ของภาพจริงเปรียบเทียบกับฟีเจอร์เวกเตอร์ของภาพที่ถูกสร้างหรือสังเคราะห์ขึ้น ค่า FID ยิ่งน้อยยิ่งบ่งบอกถึงคุณภาพของภาพที่ถูกสังเคราะห์ขึ้นมาว่าสูงกว่าหรือเทียบเท่ากันกับคุณภาพของภาพจริง ภาพที่ 3 ถึง 5 แสดงผลลัพธ์ของ FID เปรียบเทียบระหว่างชุดข้อมูลที่ใช้ฝึกสอนโมเดลตั้งต้นกับชุดข้อมูลภาพสังเคราะห์สไตล์ DOTA2 ที่งานวิจัยนี้สร้างขึ้นมา จากภาพจะเห็นว่าค่าคะแนน FID ของแต่ละโมเดลนั้นมีช่วงค่า FID ต่ำสุดแตกต่างกันไป โดยโมเดล FFHQ มีค่า FID ต่ำสุดในช่วงการฝึกสอนรอบที่ 5,000 ในขณะที่โมเดล LSUN Church และโมเดล LSUN Horse มีค่า FID ต่ำสุดที่การฝึกสอนรอบที่ 12,500 และ 7,500 ตามลำดับ

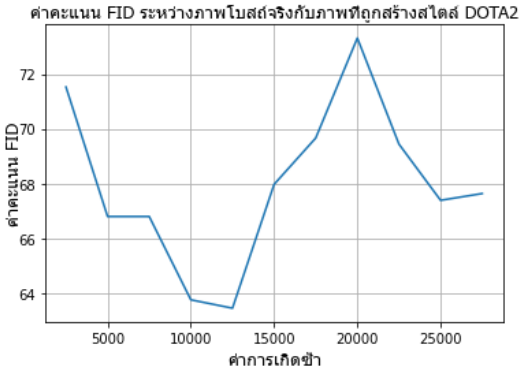
การเปรียบเทียบเชิงคุณภาพ (Qualitative comparison)

สำหรับการเปรียบเทียบเชิงคุณภาพ ผู้วิจัยได้นำภาพที่ถูกสร้างขึ้นจากทั้ง 3 โมเดลไปให้ผู้เล่น DOTA2 ประเทศไทยจำนวน 12 ท่านประเมิน โดยผู้เล่นทั้ง 12 ท่านนี้มีประสบการณ์การเล่น DOTA2 อย่างต่ำหนึ่งปีขึ้นไป และอยู่ในกลุ่ม Facebook ชื่อ DOTA2 Thailand ซึ่งเป็นกลุ่มที่รวบรวมผู้เล่นที่มีความชอบและสนใจในเกม DOTA2 ทั้งนี้ในการประเมินผู้เข้าร่วมการประเมินจะได้รับชมภาพจากแต่ละโมเดล โมเดลละ 25 ภาพ รวมทั้งหมด 75 ภาพ โดยไม่ทราบว่าภาพนั้น ๆ ถูกสร้างขึ้นจากโมเดลใด จากนั้นจะมีคำถามทั้งหมด 6 คำถามเพื่อให้ผู้เข้าร่วมการประเมินตอบเพื่อพิจารณาภาพที่ถูกสร้างขึ้นในหัวข้อต่าง ๆ

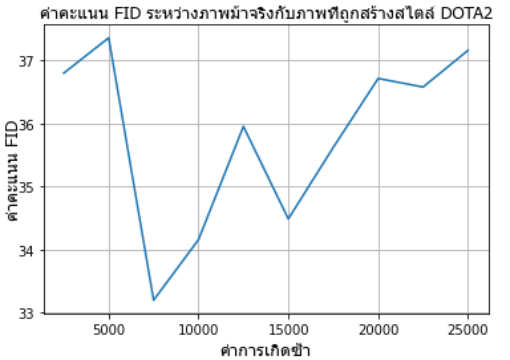
โดยผลสรุปคะแนนเฉลี่ยในด้านความสวยงาม (ภาพที่ 6) พบว่าโมเดล LSUN Church ได้คะแนนเฉลี่ยสูงสุด 7.17 ตามด้วยโมเดล FFHQ ที่คะแนน 7.08 และ โมเดล LSUN Horse ที่คะแนน 6.67 คะแนนตามลำดับ ส่วนในด้านความคมชัด (ภาพที่ 7) พบว่าผลคะแนนเฉลี่ยเรียงลำดับจากมากไปน้อยคือ โมเดล LSUN Church ได้ 8.17 โมเดล FFHQ ได้ 8.08 และโมเดล LSUN Horse ได้ 7.83 สุดท้ายคะแนนเฉลี่ยด้านความมีกลิ่นไอ DOTA2 (ภาพที่ 8) พบว่า โมเดล FFHQ ได้คะแนนเฉลี่ยสูงสุด 8.33 ตามด้วยโมเดล LSUN Church ที่คะแนน 7.42 และ โมเดล LSUN Horse ที่คะแนน 7.33 ตามลำดับ



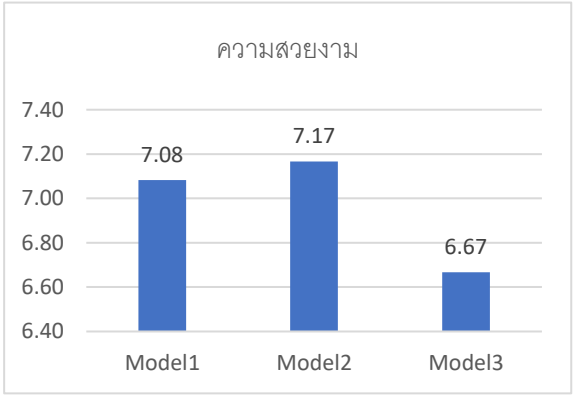
รูปที่ 3 ค่าคะแนน FID (แกนตั้ง) ที่การฝึกสอนรอบต่าง ๆ (แกนนอน) เปรียบเทียบระหว่างภาพหน้ามนุษย์จริงกับภาพที่ถูกสร้างขึ้นในสไตล์ DOTA2



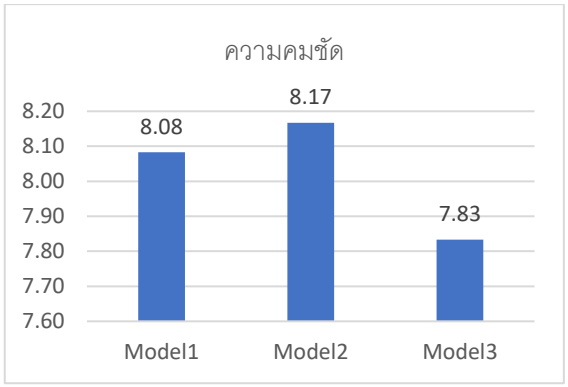
รูปที่ 4 ค่าคะแนน FID (แกนตั้ง) ที่การฝึกสอนรอบต่าง ๆ (แกนนอน) เปรียบเทียบระหว่างภาพโบสถ์จริงกับภาพที่ถูกสร้างขึ้นในสไตล์ DOTA2



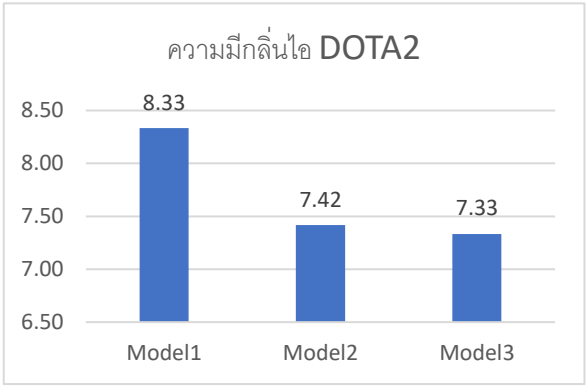
รูปที่ 5 ค่าคะแนน FID (แกนตั้ง) ที่การฝึกสอนรอบต่าง ๆ (แกนนอน) เปรียบเทียบระหว่างภาพม้าจริงกับภาพที่ถูกสร้างขึ้นในสไตล์ DOTA2



รูปที่ 6 ค่าเฉลี่ยคะแนนความสวยงาม (คะแนนเต็ม 10) ของภาพตัวละคร DOTA2 ที่สร้างขึ้นที่สร้างขึ้นโดยโมเดล FFHQ (Model1) โมเดล LSUN Church (Model2) และโมเดล LSUN Horse (Model3)

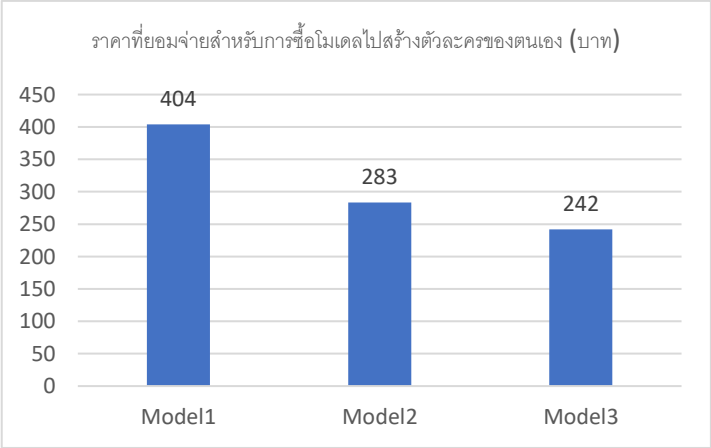


รูปที่ 7 ค่าเฉลี่ยคะแนนความคมชัด (คะแนนเต็ม 10) ของภาพตัวละคร DOTA2 ที่สร้างขึ้นที่สร้างขึ้นโดยโมเดล FFHQ (Model1) โมเดล LSUN Church (Model2) และโมเดล LSUN Horse (Model3)

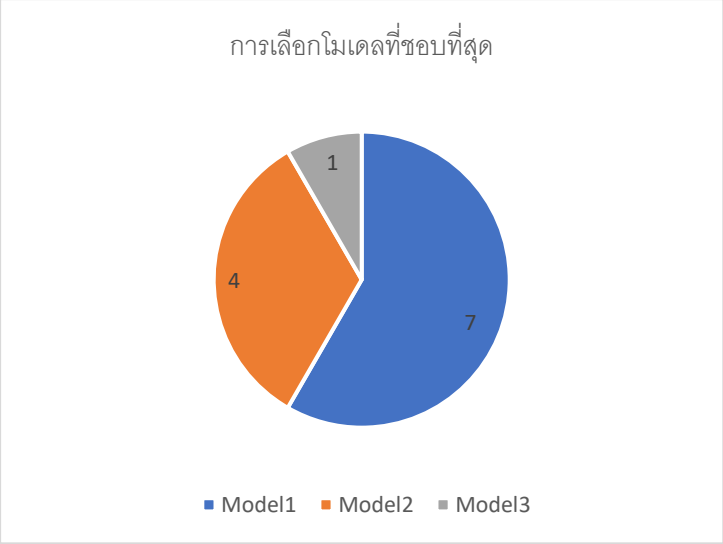


รูปที่ 8 ค่าเฉลี่ยของคะแนนความมีกลิ่นไอ DOTA2 (คะแนนเต็ม 10) ของภาพตัวละคร DOTA2 ที่สร้างขึ้นที่สร้างขึ้นโดยโมเดล FFHQ (Model1) โมเดล LSUN Church (Model2) และโมเดล LSUN Horse (Model3)

รูปที่ 9 ตัวอย่างภาพตัวละครที่ถูกสร้างขึ้นใหม่จากโมเดล FFHQ ที่มีความใกล้เคียงกับตัวละคร Grimstroke และ Axe ที่มีอยู่แล้วในเกม DOTA2



รูปที่ 10 ราคาที่ยอมจ่ายสำหรับการซื้อโมเดลเพื่อไปใช้สร้างตัวละครของตนเอง สำหรับโมเดล FFHQ (Model1) โมเดล LSUN Church (Model2) และโมเดล LSUN Horse (Model3)



รูปที่ 11 การเลือกโมเดลที่ชอบที่สุดในภาพรวมระหว่างโมเดล FFHQ (Model1) โมเดล LSUN Church (Model2) และโมเดล LSUN Horse (Model3)

ทั้งนี้จากการสัมภาษณ์เพิ่มเติมพบว่าผู้เข้าร่วมการทดสอบส่วนใหญ่เห็นตรงกันว่าเมื่อได้เห็นภาพที่ถูกสร้างขึ้นจากโมเดล FFHQ สามารถรู้ได้ในทันทีว่าเป็นภาพจากเกม DOTA2 เช่น แม้จะเป็นภาพของตัวละครใหม่ที่ผู้เข้าร่วมการทดลองไม่รู้จักมาก่อน อย่างตัวละครใหม่สีแดงที่ถูกสร้างขึ้นในภาพที่ 9 (ซ้าย) ผู้เข้าร่วมการทดสอบจำนวน 7 จาก 12 ท่านก็ยังให้ความเห็นระบุลงมาได้ว่าตัวละครใหม่นี้มีความใกล้เคียงกันกับตัวละครชื่อ Grimstroke และ Axe ซึ่งเป็นตัวละครที่มีอยู่แล้วในเกม DOTA2

ในส่วนของราคาที่ผู้เข้าร่วมการทดลองจะยอมจ่ายเพื่อซื้อโมเดลไปสร้างตัวละครของตนเองนั้น พบว่าโมเดล FFHQ มีราคาที่ผู้เข้าร่วมการทดสอบยอมจ่ายเฉลี่ยสูงที่สุด 404 บาท ตามมาด้วยโมเดล LSUN Church ในราคา 283 บาท และโมเดล LSUN Horse ในราคา 242 บาท แสดงดังภาพที่ 10 ทั้งนี้จากการสอบถามเหตุผลเพิ่มเติมผู้เข้าร่วมการทดลองแจ้งว่าต้องการนำโมเดลไปใช้สร้างภาพตัวละคร DOTA2 จากใบหน้าของตัวเอง หรือนำไปใช้สร้างภาพวาดจากผู้ชม (Fanart) เสียเป็นส่วนใหญ่

สุดท้ายในการประเมินโดยภาพรวมว่าภาพของโมเดลใดที่ผู้เข้าร่วมการทดลองชอบมากที่สุด พบว่าโมเดล FFHQ มีผู้เข้าร่วมการทดสอบชอบมากที่สุดทั้งหมด 7 คน ตามด้วยโมเดล LSUN Church จำนวน 4 คน และ โมเดล LSUN Horse จำนวน 1 คนตามลำดับ โดยปัจจัยหลักที่ตัดสินใจเลือกมาจากความมีกลิ่นไอของ DOTA2 เป็นสำคัญ ดังภาพที่ 11 ทั้งนี้ในการทดลองครั้งนี้ผู้วิจัยพบว่าผู้เข้าร่วมการทดลองมีความชื่นชอบและตื่นเต้นที่จะได้สร้างตัวละคร DOTA2 ขึ้นมาใหม่ด้วยตนเอง และมีข้อเสนอแนะเพิ่มเติมให้แก่ผู้วิจัย คือ การเพิ่มตัวเลือกในการระบุเพศและเผ่าพันธุ์ของตัวละครที่จะสร้างขึ้นได้ เช่น ตัวละครเพศหญิงเผ่าเอลฟ์ หรือ ตัวละครเพศชายเผ่าสัตว์ป่า เป็นต้น

5. สรุปผลการศึกษา

งานวิจัยนี้นำเครือข่ายขัดแย้งกำเนิดชื่อ StyleGANv2 ที่ถูกฝึกสอนมาด้วยชุดข้อมูล 3 ชุดซึ่งแตกต่างกันมาทำการฝึกสอนต่อยอดบนชุดข้อมูลภาพฮีโร่ของเกม DOTA2 ด้วยเทคนิคการสร้างภาพช็อตจำนวนน้อยผ่านการโต้ตอบข้ามโดเมน ผลการเปรียบเทียบโมเดลทั้งสามที่ทำการทดลองพบว่าโมเดลทุกตัวสามารถสร้างภาพสีเข้มและสดที่เป็นสไตล์ของภาพตัวละครใน DOTA2 ได้ แต่ในแง่ความถูกต้องนั้นโมเดลที่ประยุกต์มาจากโมเดล FFHQ สามารถดัดแปลงส่วนตาและจมูกของคน มาเป็นตาและจมูกของฮีโร่ได้อย่างถูกต้องที่สุด แต่ส่วนปากจะถูกลบออกไป ผลการประเมินจากผู้เล่นเกม DOTA2 พบว่าโมเดลที่ต่อยอดจาก FFHQ เป็นโมเดลที่มีผู้เข้าร่วมการทดลองชื่นชอบและให้ราคาในการซื้อมากที่สุด โดยปัจจัยหลักที่ตัดสินใจเลือกมาจากความสามารถในการสร้างภาพที่มีกลิ่นไอของ DOTA2 เป็นสำคัญ อีกทั้งผู้เข้าร่วมการทดลองแสดงความสนใจในโมเดลลักษณะนี้ที่ทำให้ผู้เล่นเกมสามารถสร้างและออกแบบตัวละครในเกมได้เอง

สำหรับการทำวิจัยต่อยอดในอนาคตนั้น ผู้วิจัยคาดว่าจะทำการทดลองเพื่อหาจำนวนของรูปภาพฝึกสอนที่เหมาะสมที่สุด (Optimal number of training images) ว่าควรเป็นจำนวนเท่าไหร่จึงจะได้ผลลัพธ์ที่สมดุลที่สุดในปัจจัยประเมินด้านต่าง ๆ นอกจากนี้ผู้วิจัยยังต้องการจะเพิ่มจำนวนของผู้เข้าร่วมการทดลองให้มากขึ้นเพื่อให้ผล

การประเมินเชิงคุณภาพสะท้อนความคิดเห็นที่แท้จริงของผู้เล่นเกม DOTA2 ส่วนใหญ่ในประเทศไทยได้ รวมถึงอาจมีการใช้เทคนิควิธีการวิจัยเชิงคุณภาพอื่น ๆ เพิ่มเติมนอกจากเหนือจากแบบสอบถามและการสัมภาษณ์ออนไลน์ เช่น การสัมภาษณ์เชิงลึก การอภิปรายกลุ่มย่อย เป็นต้น (Creswell & Creswell, 2018) สุดท้ายคือการทดลองใช้เจนเนอเรทีฟโมเดลตัวอื่นที่นอกเหนือจากเครือข่ายขัดแย้งกำเนิดเพื่อเปรียบเทียบผลการทดลอง

กิตติกรรมประกาศ

กลุ่มผู้วิจัยขอขอบพระคุณผู้เข้าร่วมการทดลองจาก Facebook กลุ่ม DOTA2 THAILAND ทั้ง 12 ท่าน เป็นอย่างสูงที่ได้สละเวลาเข้าร่วมทดสอบ แสดงความคิดเห็น และ ให้ข้อเสนอแนะที่มีคุณค่าเพื่อปรับปรุงผลลัพธ์ของงานวิจัยนี้ให้ดียิ่งขึ้นต่อไปในอนาคต

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*. https://doi.org/10.14414/jebav.v22i3.1873

Auc, T., (2013). *DOTA2 Hero* [Photograph]. Retrieved from https://auct.eu/dota2-hero-images

Back, J. (2021). *Fine-Tuning StyleGAN2 For Cartoon Face Generation*. https://arxiv.org/abs/2106.12445

Clement, J. (2021). *Number of monthly active users (MAU) of DOTA 2 worldwide as of February 2021*. 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)*, 2672-2680.

Grand, V. R. (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*, *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://arxiv.org/abs/2004.08066

Heusel, M., Ramsauer, H., Unterthiner, T., & Nessler, B. (2017). *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*. 6629-6640.

Jin, Y., Zhang, J., Li, M., Tian, Y., Zhu, H., & Fang, Z. (2017). *Towards the Automatic Anime Characters Creation with Generative Adversarial Networks. NIPS 2017 Workshop on Machine Learning for Creativity and Design*. https://nips2017creativity.github.io/doc/High_Quality_Anime.pdf

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Timo, A. (2020). *Analyzing and Improving the Image Quality of StyleGAN. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2020)*. https://doi.org/10.1109/CVPR42600.2020.00813

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.

Kynkäänniemi, T., Karras, T., Aittala, M., Aila, T., & Lehtinen, J. (2022). *The Role of ImageNet Classes in Fréchet Inception Distance*. *arXiv preprint arXiv:2203.06026*

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*, *24*, 4077-4091. https://doi.org/10.1109/tmm.2021.3113786

Ojha, U., Li, Y., Lu, J., Efros, A. A., Lee, Y. J., Shechtman, E., & Zhang, R. (2021). *Few-shot image generation via cross-domain correspondence*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10743-10752.

Park, B.-W. & Lee, Kun, C. (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. Y., Jiang, L., Liu, C., Yang, M. H., & Yang, W. (2021). *Regularizing generative adversarial networks under limited data. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (pp. 7921-7931). https://doi.org/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. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. (pp. 3892-3901). https://doi.org/10.1109/wacv51458.2022.00019