

## การเปลี่ยนทรงผมเสมือนจริง: การกำหนดตำแหน่งใบหน้าและผมด้วยแผนที่ความหมายสำหรับ SDEdit

### Realistic Hairstyle Try-On: Face and Hair Image Mapping Using Semantic Maps for SDEdit

สรยุทธ มีอิม, ภาพพัทธ์ เต็กตระกูล, ด้ลพงษ์ อรรคนุตร, วีรพล จิรจรีต

ภาควิชาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี sorayut.mild@mail.kmutt.ac.th

#### บทคัดย่อ

การสร้างภาพด้วยการเรียนรู้ของเครื่องสามารถเปลี่ยนสีผมหรือทรงผมของบุคคลในภาพได้ บทความนี้นำเสนอวิธีการสร้างและตัดต่อภาพบุคคลกับทรงผมที่ต้องการ โดยใช้แผนที่ความหมาย จากนั้นจึงทำการกำหนดตำแหน่งใบหน้าและผม สุดท้ายภาพจะถูกเติมเต็มโดยใช้วิธี Fast Marching Method และ Stochastic Differential Editing (SDEdit) ผลการทดลองแสดงให้เห็นว่าวิธีที่นำเสนอสามารถควบคุมได้ทั้งรูปร่างของทรงผมและสีผมโดยใช้รูปภาพบุคคลที่อยากเปลี่ยนทรงผมและรูปภาพบุคคลที่มีทรงผมที่ต้องการเพียงรูปเดียว นอกจากนี้ยังสามารถเปลี่ยนทรงผมได้ในกรณีที่ทรงผมเดิมของรูปภาพบุคคลถูกบดบังด้วยทรงผมเก่า

**คำสำคัญ:** การเปลี่ยนทรงผมเสมือนจริง แผนที่ความหมาย SDEdit

#### Abstract

Machine learning-based image generation can create new person face images with new hair colors or hairstyles. This paper presents synthesis and editing method to modify hairstyles in the images by semantic maps. The face and hair images are mapped and inpainted using fast marching method and Stochastic Differential Editing (SDEdit). The experimental results shows that the proposed method controls both hairstyles and color effectively with single target hairstyle image. Moreover, the method is able to generate hairstyles in case of occluded face images.

**Keywords:** Realistic hairstyle try-on, Semantic maps, SDEdit

#### 1. บทนำ

การเปลี่ยนทรงผมของบุคคลในรูปภาพ โดยสามารถควบคุมทรงผมที่ต้องการเปลี่ยนได้ และไม่ทำให้ส่วนอื่นของใบหน้าเปลี่ยนไป สามารถนำไปใช้เพื่อประกอบการตัดสินใจในการเปลี่ยนทรงผม แต่การเปลี่ยนทรงผมของบุคคลในรูปภาพนั้นเป็นงานที่ท้าทาย อันเนื่องมาจากความหลากหลายของทรงผมทั้งในด้านของรูปร่าง หรือสีผม และสังเกตเห็นความผิดพลาดได้ง่าย แตกต่างจากส่วนอื่นของใบหน้ามนุษย์ นอกจากนั้นความท้าทายอีกอย่าง คือ การเปลี่ยนทรงผมแต่ยังคงรักษาสีผมเดิม

รูปภาพใบหน้าให้ยังคงเดิม และยังคงความท้าทายจากส่วนที่รูปภาพไม่สมบูรณ์ (Artifacts) ที่เกิดขึ้นในรูปภาพ โดยเฉพาะการเปลี่ยนทรงผมจากคนที่ไม่มีทรงผมบดบังส่วนอื่นของใบหน้าเป็นทรงผมที่ไม่บดบังส่วนของใบหน้า

บทความนี้นำเสนอ Face and Hair Image Mapping Using Semantic Maps for SDEdit Conditional Hair Changing Image วิธีการสร้างรูปภาพที่เป็นการตัดต่อรูปทรงผมโดยอัตโนมัติของรูปภาพบุคคลที่เป็นเป้าหมายในการเปลี่ยนทรงผมหรือรูปภาพเป้าหมาย (Target image) และรูปภาพของบุคคลที่มีทรงผมที่ต้องการ (Source image) โดยใช้วิธีการ Semantic segmentation แต่ละส่วนของรูปภาพ จากนั้นนำรูปภาพที่เป็นการตัดต่อไปเข้า SDEdit ซึ่งเป็น Generative model เพื่อให้รูปภาพดูเรียบเนียนสมจริงมากยิ่งขึ้น โดยรูปภาพผลลัพธ์ที่ได้ เราสามารถกำหนดได้ทั้งรูปร่างของทรงผมและสีผมของผม โดยไม่ทำให้รูปภาพส่วนอื่นของใบหน้าผิดเพี้ยน

วัตถุประสงค์ของงานวิจัยนี้ คือ ต้องการสร้างโปรแกรมสำหรับใช้งานเพื่อประกอบการตัดสินใจในการเปลี่ยนทรงผม โดยให้ได้รูปภาพผลลัพธ์เสมือนจริงมากที่สุด

#### 2. ทฤษฎีที่เกี่ยวข้อง

##### 2.1. Image Segmentation for Hair Editing

งานด้านการเปลี่ยนทรงผมพัฒนาขึ้นอย่างรวดเร็วในช่วงหลัง อันเนื่องมาจากการพัฒนาขึ้นอย่างรวดเร็วของการเรียนรู้เชิงลึกชนิด Generative model และได้มีการนำ Image segmentation มาทำการแยกส่วนของใบหน้า เพื่อช่วยในการเลือกส่วนที่จะเปลี่ยนแปลงได้ ตัวอย่างเช่น บทความที่นำเสนอวิธีการสร้างรูปภาพใบหน้าที่สามารถกำหนดได้โดยใช้ Semantic maps ซึ่งสามารถเปลี่ยนรูปร่างของทรงผมได้ [1] แต่ไม่สามารถเปลี่ยนสีผมได้ และการเปลี่ยนทรงผมอาจจะส่งผลให้ส่วนอื่นของใบหน้าเปลี่ยนไป หรือนำเสนอวิธีการ Segmentation ผมเพื่อเปลี่ยนสีผม [2] แต่ไม่สามารถเปลี่ยนรูปร่างของทรงผมได้

##### 2.2. Fast Marching Method

Fast Marching Method [3] โดยหลักการทำงานเบื้องต้น คือ เริ่มกระบวนการจากบริเวณขอบภายนอกสุด ของส่วนที่รูปภาพไม่สมบูรณ์ (Artifacts) โดยจะทำการหาค่าของฟังก์ชันในแต่ละจุดจากค่าเฉลี่ยถ่วง

นำหน้าจากพิกเซลรอบข้างด้านนอกของขอบ ตามรัศมีที่กำหนดเข้าสู่ศูนย์กลางของส่วนที่รูปภาพไม่สมบูรณ์

### 2.3. SDEdit

Stochastic Differential Editing (SDEdit) [4] คือ Generative model ที่ใช้ Stochastic differential equation (SDEs) โดยข้อมูลขาเข้าจะเป็นรูปภาพที่ถูกตัดต่อ โดยเริ่มต้นจะทำการใส่ noise ลงไปในรูปภาพตาม SDE และทำการ Denoise โดยใช้การจำลอง SDE ย้อนกลับ ซึ่งข้อดีของ SDEdit ที่เหนือกว่า Generative model อื่น ๆ ที่ใช้สำหรับงานรูปภาพในปัจจุบัน คือ ไม่จำเป็นต้องใช้ loss function ที่ถูกออกแบบมาเฉพาะบางงาน ซึ่งตามปกติถ้าเป็น GAN inversion [5] จะต้องทำ และ SDEdit ไม่จำเป็นต้องเก็บชุดข้อมูลใหม่เป็นสองชุด คือ รูปภาพต้นฉบับ และรูปภาพแบบที่ต้องการ ซึ่งจำเป็นสำหรับ Conditional GANs [6, 7] นอกจากนั้น ยังไม่ต้องทำการฝึกสอนใหม่เมื่อต้องการเปลี่ยนการใช้งาน เช่น ต้องการเปลี่ยนจากทรงผม เป็นดวงตา ก็ไม่จำเป็นต้องฝึกสอนใหม่

## 3. วิธีการดำเนินงานวิจัย



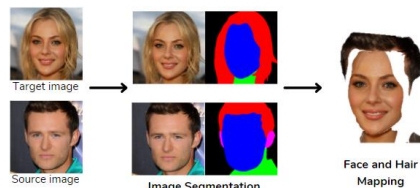
รูปที่ 1 ภาพรวมขั้นตอนการทำงานของวิธีการทำงานของงานวิจัย

จากรูปที่ 1 แสดงภาพรวมขั้นตอนการทำงานของวิธีการทำงานของงานวิจัย โดยแบ่งเป็น 4 ส่วน: Image segmentation โดยใช้ U-NET [8], Face and Hair Mapping และ Image Inpainting ที่ใช้ Fast Marching Method [3] และ SDEdit [4] โดยอธิบายการทำงานในแต่ละส่วนดังต่อไปนี้

### 3.1. Image Segmentation

เนื่องจากวิธีการทำงานวิจัยนี้เสนอวิธีที่ใช้ Semantic maps เพื่อเป็นการแบ่งรูปภาพออกเป็น 5 ส่วน ได้แก่ ผม, ใบหน้า, หู, คอกับเสื้อ และพื้นหลัง เพื่อนำไปใช้ในการตัดต่อรูป โดยใช้ U-NET [8] ซึ่งเป็นสถาปัตยกรรม Convolutional neural network ถูกนำมาใช้ในงานด้าน Image segmentation สำหรับการฝึกสอนมีรูปภาพที่ใช้มี 2 ชุด คือ ชุดรูปภาพต้นฉบับและชุดรูปภาพที่เป็น Semantic maps ของรูปภาพต้นฉบับนั้น สำหรับการทดลองนี้ชุดข้อมูลที่ใช้คือ CelebAMask-HQ ซึ่งเป็นชุดข้อมูลที่รูปภาพที่เป็นใบหน้านุคคลและรูปภาพ Semantic maps ของใบหน้านุคคลนั้น ซึ่งถูกนำเสนอโดย MaskGAN [1] สำหรับการใช้งานรูปภาพขาเข้าจะเป็นรูปภาพเป้าหมาย (Target image) ซึ่งเป็นรูปภาพของบุคคลที่เราต้องการเปลี่ยนทรงผม และรูปภาพของบุคคลที่มีทรงผมที่ต้องการ (Source image)

### 3.2. Face and Hair Mapping



รูปที่ 2 ตัวอย่างรูปภาพขาเข้า Face and hair mapping และรูปภาพผลลัพธ์

สำหรับขั้นตอนนี้ ต้องการผลลัพธ์ที่เป็น รูปภาพทรงผมของรูปภาพของบุคคลที่มีทรงผมที่ต้องการ (Source image) รวมอยู่กับใบหน้าของรูปภาพเป้าหมาย (Target image) ซึ่งขั้นตอนนี้นำเสนอวิธีการในการ Mapping ทั้งสองรูปภาพเข้าหากันได้เพื่อให้ได้ผลลัพธ์ตามรูปที่ 2 โดยที่ข้อมูลขาเข้าที่จำเป็นต้องใช้ในวิธีการต่าง ๆ คือ รูปภาพของบุคคลที่มีทรงผมที่ต้องการ (Source image), รูปภาพเป้าหมาย (Target image) รวมถึงรูปภาพของผลลัพธ์ของขั้นตอน Image segmentation ของทั้ง Source image และ Target image ทั้งหมด 4 รูปภาพ ซึ่งทุกรูปภาพจะต้องมีขนาดที่เท่ากันเท่านั้น โดยใช้วิธีการต่าง ๆ ได้แก่ Image face landmark, Image transform, Image cropping และ Image replacing ตามลำดับ ซึ่งจะมีวิธีการดำเนินการต่าง ๆ ดังนี้

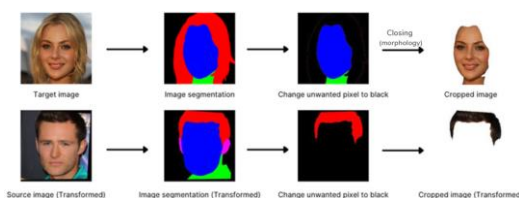
#### 3.2.1. Image face landmark

Image face landmark คือ การหาตำแหน่งต่าง ๆ ตามที่ต้องการบนใบหน้าของรูปภาพ โดยใช้ MediaPipe Face Mesh [9] โดยงานวิจัยนี้ได้ใช้กระบวนการดังกล่าวในการหาตำแหน่งบนรูปภาพของ Target image และ Source image จำนวน 4 ตำแหน่ง ได้แก่ หน้าผาก, ตาซ้าย, ตาขวา และ คาง

#### 3.2.2. Image transform

Image transform คือ วิธีการ Transform Source image ให้ตำแหน่งผลลัพธ์ทั้ง 4 จุด ของวิธีการที่ 1 (Image face landmark) ให้ได้ตำแหน่งเดียวกันกับ Target image ซึ่งงานวิจัยนี้ได้นำเสนอวิธีในการ Transform โดยใช้ Perspective transformation เพื่อให้ได้ผลลัพธ์ที่ตำแหน่งทรงผมของ Source image อยู่ตรงตำแหน่งส่วนหัวของ Target image พอดี

#### 3.2.3. Image cropping

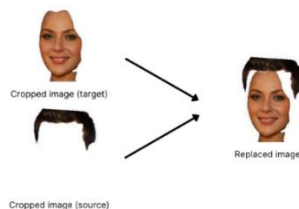


รูปที่ 3 ขั้นตอนการทำกระบวนการ Image cropping

Image cropping คือ วิธีการ Cropping รูปภาพตาม Semantic maps ที่ต้องการ ซึ่งงานวิจัยนี้ได้ทำการ Cropping Source image ที่ผ่าน วิธีการที่

2 (Image transform) ร่วมกับ Image segmentation ของภาพนั้น โดยทำการ Cropping ตาม Semantic maps ที่เป็นพื้นที่สีแดง เพื่อให้ได้ทรงผมของ Source image เท่านั้น ในส่วนของ Target image ก็จะมีการ Cropping ส่วนของเส้นผมและพื้นหลังออก ซึ่งวิธีการในการ Cropping คือเปลี่ยน Pixel สีของ Semantic maps ที่ไม่ต้องการเป็นพิกเซลสีดำแล้ว เปรียบเทียบรูปต้นฉบับกับรูป Semantic maps ที่ถูกตัดแล้ว เพื่อแทนที่ตำแหน่งของพิกเซลบนรูปภาพต้นฉบับ ด้วยตำแหน่งของพิกเซลสีดำ ใน Image segmentation ด้วยพิกเซลสีขาว และมีการใช้ Morphological transformation ชนิด Closing หลังจากมีการ Cropping ทรงผมของ Target image เพื่อลบขอบเส้นผม เป็นไปตามรูปที่ 3

### 3.2.4. Image replacing



รูปที่ 4 ตัวอย่างรูปภาพพาเข้า Image replacing และรูปภาพผลลัพธ์

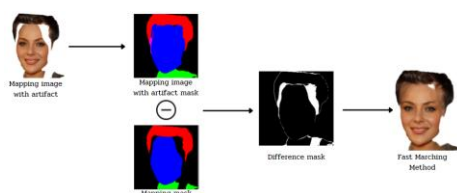
Image replacing คือ การนำภาพผลลัพธ์ของ วิธีการที่ 3 (Image cropping) ได้แก่ ทรงผมของ Source image และ Target image ที่ไม่มีทรงผม มารวมกันเป็นภาพเดียวกัน ซึ่งผลลัพธ์ที่ต้องการในวิธีการนี้คือภาพที่มีทรงผมของ Source image ทับกับ Target image ที่ถูกตัดทรงผมออก เป็นไปตามรูปที่ 4 โดยที่งานวิจัยนี้มีวิธีการ Replacing โดยการใช้ภาพทรงผมของ Source image พร้อมกับพื้นหลังสีขาวเป็นภาพตั้งต้น และทำการแทนที่พิกเซลสีขาวด้วยพิกเซลของภาพ Target image ที่ไม่มีทรงผม เพื่อให้ได้ผลลัพธ์ที่ตามต้องการ

### 3.3. Image Inpainting

ขั้นตอนนี้จะทำการเติมส่วนที่ขาดหายไปจากขั้นตอน Face and Hair Mapping โดยจะแบ่งออกเป็น 2 ส่วน คือ Fast Marching method สำหรับเติมส่วนที่ขาดหายไปโดยใช้วิธีการเฉลี่ยถ่วงน้ำหนักจากพิกเซลรอบข้าง และใช้ SDEdit สำหรับทำให้รูปภาพเรียบเนียนขึ้น

#### 3.3.1. การเติมส่วนที่ขาดหายไปของรูปภาพ โดยใช้ Fast

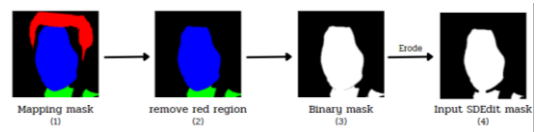
##### Marching method



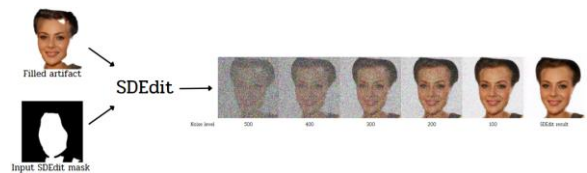
รูปที่ 5 ขั้นตอนการทำงานของขั้นตอนการเติมส่วนที่ขาดหายไปของรูปภาพ โดยใช้ Fast Marching method

กระบวนการนี้เป็นการเติมส่วนที่รูปภาพไม่สมบูรณ์ (Artifacts) ที่เกิดขึ้นหลังจากกระบวนการ Face and Hair Mapping ดังรูปที่ 4 โดยจะแบ่งเป็น 3 ขั้นตอน คือ 1. นำภาพที่ได้จากการทำ Face and Hair Mapping ไปทำ Image segmentation 2. นำภาพผลลัพธ์ที่ได้จากขั้นตอนที่ 1 และ ภาพ Mapping mask มาแปลงเป็นภาพไบนารีจากนั้นนำมาหาผลต่างระหว่างภาพทั้งสอง 3. ภาพผลลัพธ์ที่ได้จากขั้นตอนที่ 2 และ ภาพ Mapping image with artifact เข้าสู่ Fast Marching method [3] เป็นไปตามรูปที่ 5

#### 3.3.2. SDEdit



รูปที่ 6 ขั้นตอนการสร้าง mask สำหรับ SDEdit



รูปที่ 7 รูปภาพพาเข้าและรูปภาพผลลัพธ์ของ SDEdit ในแต่ละ Noise level

รูปภาพที่ได้รับจากขั้นตอน Fast Marching method อาจจะมีบางส่วนของรูปภาพดูไม่เหมือนจริง หรือเกิดส่วนที่ขาดหายไป (Artifacts) โดยส่วนนี้จะใช้ SDEdit ช่วยปรับปรุงส่วนที่เป็น Artifacts พร้อมทั้งรูปภาพดูเหมือนจริงมากขึ้น โดยรูปภาพพาเข้า SDEdit จะมี 2 รูป คือ รูปภาพที่ถูกตัดต่อ และ mask ที่เลือกส่วนที่ต้องการให้ SDEdit แก้ไขของรูปภาพนั้น โดยที่ส่วนสีขาวของ mask คือ ส่วนของรูปภาพที่ถูกตัดต่อที่จะคงเดิมไว้ (ส่วนของใบหน้า คอ และเสื้อ) และส่วนดำของ mask คือ ส่วนของรูปภาพที่ถูกตัดต่อที่จะให้ SDEdit เปลี่ยนแปลง โดยสร้าง mask ได้จากนำ mask ที่ได้จาก Face and hair mapping ลบส่วนที่เป็นผมออก (สีแดงของ mask) จากนั้นทำการเปลี่ยนเป็น binary mask และทำการ Morphology ชนิด Erosion ตามรูปที่ 6 จากนั้นนำ mask ที่ได้กับรูปภาพที่ได้รับจากขั้นตอน Fast Marching method เข้า SDEdit จะได้ผลลัพธ์เป็นไปตามรูปที่ 7

### 4. ผลการวิจัย

ในการทดลองของงานวิจัยนี้ ได้ทำการทดลองใช้ Target image และ Source image ในรูปแบบต่าง ๆ เพื่อเปลี่ยนทรงผมของบุคคลในภาพ ได้รูปภาพผลลัพธ์เป็นไปตามรูปที่ 8, รูปที่ 9, รูปที่ 10 และรูปที่ 11 โดยภาพบุคคลด้านซ้ายบน คือ Target image หรือรูปภาพบุคคลที่ต้องการเปลี่ยนทรงผม รูปภาพซ้ายล่าง คือ Source image หรือภาพของบุคคลที่มีทรงผมที่ต้องการ และรูปภาพขวามือ คือ ภาพผลลัพธ์ที่เปลี่ยนทรงผม

แล้ว และภาพขาเข้า SDEdit และภาพผลลัพธ์ในแต่ละ Noise level เป็นไปตามรูปที่ 12



รูปที่ 8 ตัวอย่างรูปภาพผลลัพธ์ของงานวิจัยในการเปลี่ยนรูปภาพของบุคคลทรงผมขาวไปเป็นทรงผมสั้น



รูปที่ 9 ตัวอย่างรูปภาพผลลัพธ์ของงานวิจัยในการเปลี่ยนรูปภาพของบุคคลทรงผมสั้นไปเป็นทรงผมยาว



รูปที่ 10 ตัวอย่างรูปภาพผลลัพธ์ของงานวิจัยในการเปลี่ยนรูปภาพของบุคคลทรงผมยาวไปเป็นทรงผมยาว



รูปที่ 11 ตัวอย่างรูปภาพผลลัพธ์ของงานวิจัยในการเปลี่ยนรูปภาพของบุคคลที่มีส่วนของทรงผมบดบังใบหน้าไปเป็นทรงผมที่ไม่บดบังใบหน้า



รูปที่ 12 การเปรียบเทียบรูปภาพขาเข้า SDEdit และรูปภาพผลลัพธ์ในแต่ละ Noise level

## 5. สรุป

งานนี้นำเสนอ Face and Hair Image Mapping Using Semantic Maps for SDEdit Conditional Hair Changing Image วิธีการสร้างรูปภาพที่เป็นการตัดต่อรูปทรงผมโดยอัตโนมัติ ซึ่งใช้ประโยชน์จาก Semantic maps เพื่อนำรูปภาพที่ได้ไปเป็นรูปภาพขาเข้าสำหรับ SDEdit และได้ผลลัพธ์เป็นรูปภาพบุคคลที่มีทรงผมตามที่ต้องการ ซึ่งข้อดีของวิธีการนี้คือ 1) เราสามารถควบคุมได้ทั้งรูปร่างของทรงผม และสีผม 2) การทำงานทั้งระบบเป็นระบบอัตโนมัติ ไม่ต้องอาศัยการช่วยงานของมนุษย์ 3) ใช้รูปภาพขาเข้าแค่ 2 รูปเท่านั้น คือ Target image กับ Source image แต่การใช้งานควรเลือกค่า Noise level ให้เหมาะสม เช่น จากรูปที่ 12 การเปรียบเทียบรูปภาพขาเข้า SDEdit และรูปภาพผลลัพธ์ในแต่ละ Noise level Noise level 1000 ทรงผมผิดเพี้ยนไปจากทรงผมที่ต้องการ และที่ Noise level 300 จะไม่สามารถแก้ไขส่วนที่ขาดไปของรูปภาพได้ ซึ่งงานวิจัยนี้สามารถพัฒนาต่อได้ โดยสามารถเปลี่ยนส่วนอื่นของรูปภาพใบหน้าได้ เช่น ตา, ปาก เป็นต้น

## เอกสารอ้างอิง

- [1] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "MaskGAN: Towards Diverse and Interactive Facial Image Manipulation," *CoRR*, vol. abs/1907.11922, 2019, [Online]. Available: <http://arxiv.org/abs/1907.11922>
- [2] A. Levinshtein, C. Chang, E. Phung, I. Kezele, W. Guo, and P. Aarabi, "Real-time deep hair matting on mobile devices," *CoRR*, vol. abs/1712.07168, 2017, [Online]. Available: <http://arxiv.org/abs/1712.07168>
- [3] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," *J. Graph. Tools*, vol. 9, Jan. 2004, doi: 10.1080/10867651.2004.10487596.
- [4] C. Meng, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, "SDEdit: Image Synthesis and Editing with Stochastic Differential Equations," *ArXiv Prepr. ArXiv210801073*, 2021.
- [5] W. Xia, Y. Zhang, Y. Yang, J.-H. Xue, B. Zhou, and M.-H. Yang, "GAN Inversion: A Survey," *CoRR*, vol. abs/2101.05278, 2021, [Online]. Available: <https://arxiv.org/abs/2101.05278>
- [6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *CoRR*, vol.

abs/1611.07004, 2016, [Online]. Available:

<http://arxiv.org/abs/1611.07004>

- [7] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” *CoRR*, vol. abs/1703.10593, 2017, [Online]. Available: <http://arxiv.org/abs/1703.10593>
- [8] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015, pp. 234–241.
- [9] I. Grishchenko, A. Ablavatski, Y. Kartynnik, K. Raveendran, and M. Grundmann, “Attention Mesh: High-fidelity Face Mesh Prediction in Real-time,” *CoRR*, vol. abs/2006.10962, 2020, [Online]. Available: <https://arxiv.org/abs/2006.10962>