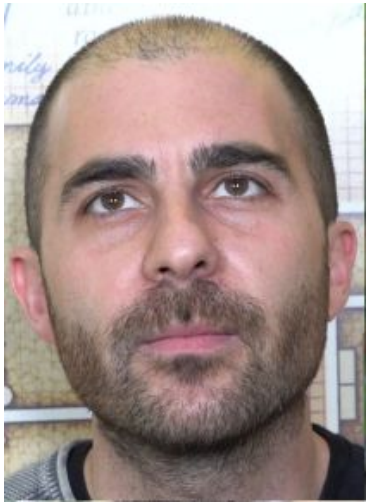
Exploring Methods Used in Face Swapping

Joshua Eklund
eklun124@umn.edu
Division of Science and Mathematics,
University of Minnesota, Morris
Morris, Minnesota, USA

Introduction: What is Face Swapping?



Source

Target

Face Swapped Image

Introduction: Why Do We Care?

- Entertainment Industry
  - “Revive” actors
    - *Rogue One: A Star Wars Story*
      - Leia Organa (Carrie Fisher)
      - Grand Moff Tarkin (Peter Cushing)
  - Snapchat
- Concerns
  - Privacy
  - Misinformation

Outline

- **Background**
  - Facial Landmark Identification and Face Segmentation
  - Image Blending
    - Multi-band blending
    - Poisson blending
- Face Swapping Approaches
  - Encoder-Decoders
  - Linear 3D Morphable Models
- Comparisons/Conclusions

Background: Identifying Facial Landmarks and Face Segmentation

- Detect faces in images
  - Facial landmark detection
    - Facial Landmarks
      - Eyes, nose, lips, etc.
    - Set of points representing landmarks
- Separate face from background or occlusions
  - Face segmentation
    - Determines visible portions of a face

Background: Image Blending

Target

Source

Background: Image Blending

Target

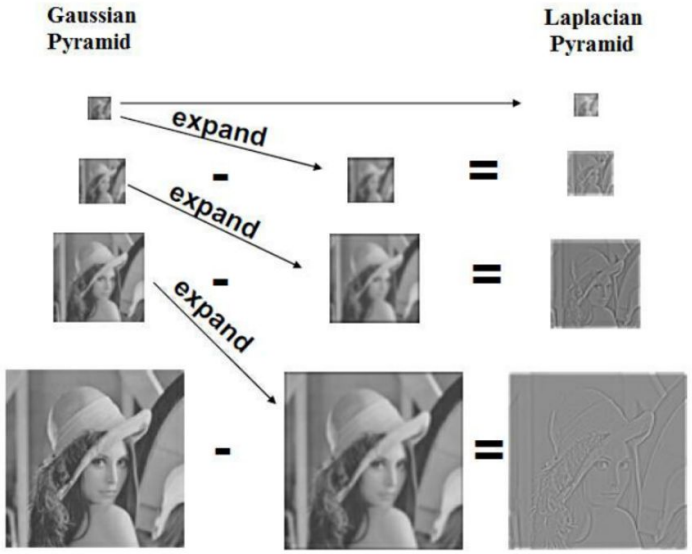
Source

Background: Image Blending

Non-blended

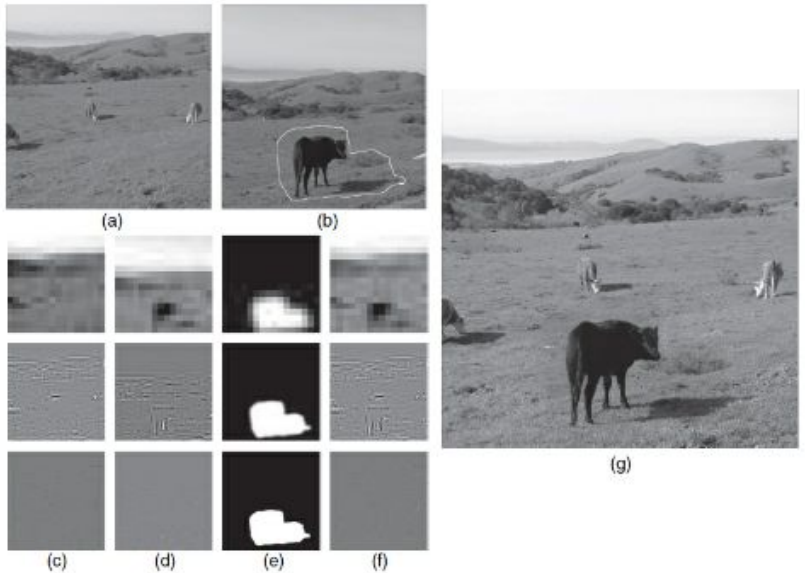
Background: Multi-band Blending

- Utilizes Gaussian and Laplacian pyramids
  - Gaussian pyramid
    - Hierarchy of images
      - Blurred and reduced in size
  - Laplacian pyramid
    - Laplacian
      - Captures edges
    - Constructed using Gaussian pyramid
      - $L_i = G_i - (G_{i+1})_{\uparrow 2}$
    - Original image
      - Reconstructed from Laplacian
        - $I = \sum L_{i\uparrow}$



Background: Multi-band Blending

- Compositie source into target
  - Gaussian pyramid G
    - Mask of source object
  - Laplacian pyramids
    - Target image L^T and Source image L^S
    - Composite image L^I
      - $L^I_i = G_i L^S_i + (1 - G_i) L^T_i$
  - Composited image
    - Constructed from L^I



Background: Multi-band Blending

Non-blended

Multi-band blending

Background: Poisson Blending

- Resolves the color/lighting mismatch
  - Gradient domain
    - Image Gradient
      - Directional change in lighting or color
- Composite source into target
  - Ω: Region of source image *S* to copy
  - ∂Ω: Boundary of Ω
  - *C*: Composited image

$$\min_{C(x,y)\in\Omega} \iint_{\Omega} \|\nabla C(x,y) - \nabla S(x,y)\|^2 \, dx \, dy$$

$$\text{s.t. } C(x,y) = T(x,y) \text{ on } \partial\Omega$$

Background: Poisson Blending

Multi-band blending

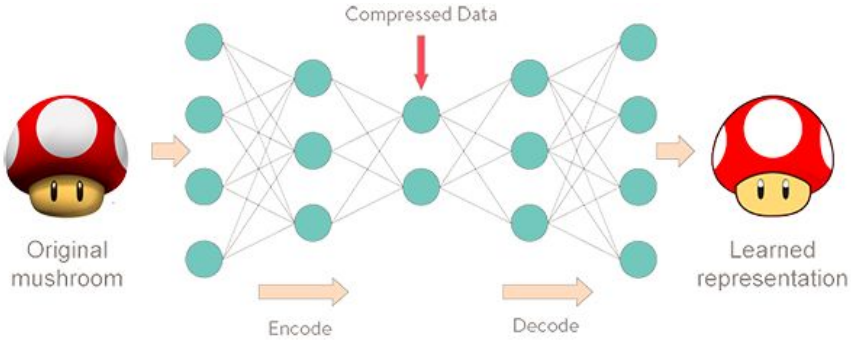
Poisson blending

Outline

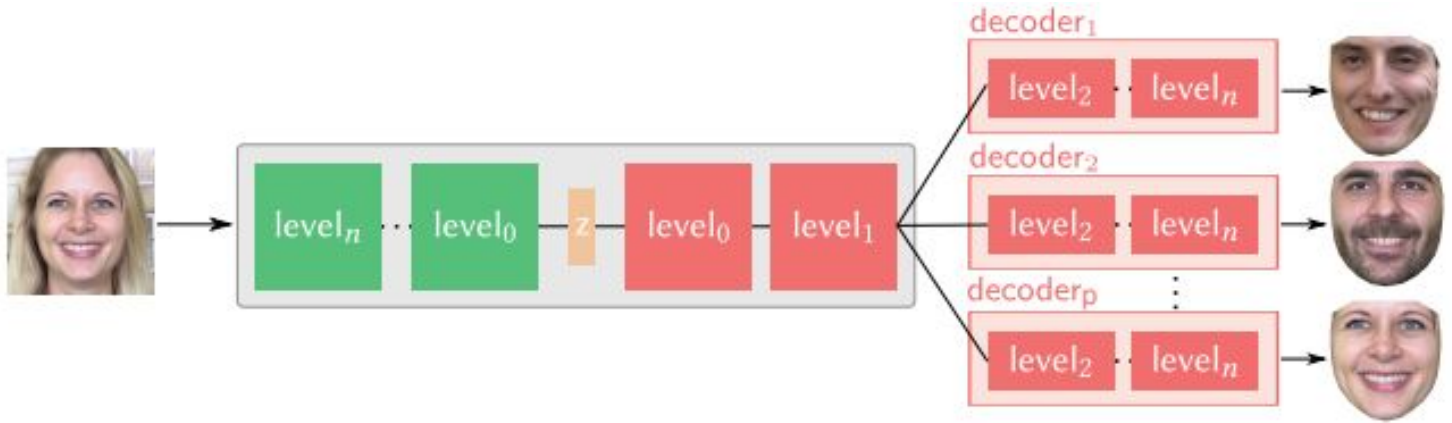
- Background
  - Facial Landmark Identification and Face Segmentation
  - Image Blending
    - Multi-band blending
    - Poisson blending
- **Face Swapping Approaches**
  - Encoder-Decoders
  - Linear 3D Morphable Models
- Comparisons/Conclusions

Face Swapping Approaches: Encoder-Decoders

- Generative model
  - “Learn” to reconstruct input data
- Consist of two parts
  - Encoder
    - Compresses data
  - Decoder
    - Reconstructs data
- Reconstruction error
  - Difference between original data and reconstruction

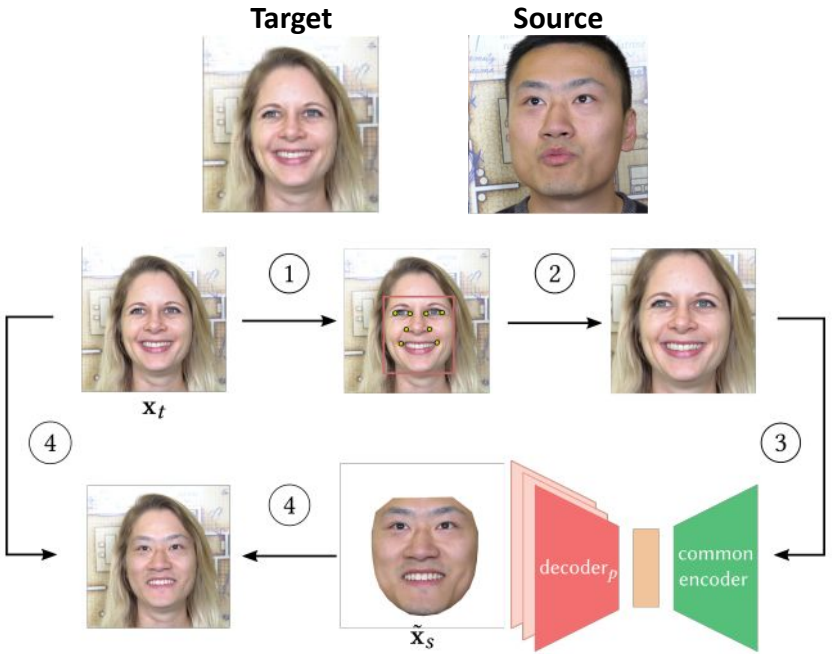


Face Swapping Approaches: Encoder-Decoders



Face Swapping Approaches: Encoder-Decoders

- Start with target
- Detect facial landmarks and normalize face (1024x1024 resolution)
- Feed face into model
- Decode face using *pth* decoder
- Blend source face into target
  - Modified multi-band blending
    - Global Contrast Factor (GCF)



Face Swapping Approaches: Linear 3D Morphable Models

- Generative Model
  - Generate 3D representation of any given face
- Vector space of shapes and textures
  - Shape vector *S*
  - Texture vector *T*
- Generation of new faces
  - Linear combinations of *S* and *T*

$$\mathbf{S} = \sum_i \alpha_i \boldsymbol{S}_i = \alpha_1 \cdot + \alpha_2 + \alpha_3 + \alpha_4 + \ldots$$

$$\boldsymbol{T} = \sum_i \beta_i \boldsymbol{T}_i = \beta_1 \cdot + \beta_2 + \beta_3 + \beta_4 + \ldots$$

Face Swapping Approaches: Linear 3D Morphable Models

- Generate 3D shape representations
  - Modify 3D shapes to match pose and expression
- Perform face segmentation
- Project 3D shape of source onto source image
  - Assign 3D vertices to 2D segmentation mask
  - Sample source image
- Transfer sampled intensities to 3D shape of target
  - Use segmentation mask of target to mask rendered intensities
- Blend rendered source face into target image
  - Poisson blending

Outline

- Background
  - Facial Landmark Identification and Face Segmentation
  - Image Blending
    - Multi-band blending
    - Poisson blending
- Face Swapping Approaches
  - Encoder-Decoders
  - Linear 3D Morphable Models
- **Comparisons/Conclusions**

Comparisons/Conclusions

- Four pairs of face swaps
  - Encoder-decoder approach
  - Linear 3D morphable model approach
- Observations
  - Linear 3D morphable approach
    - Struggles to match pose and expression of target face
      - Noticeable around eyes and mouth
  - Blending
    - Encoder-decoder is superior
      - Significant difference in lighting

Comparisons/Conclusions

	Encoder-Decoder	**Linear 3D Morphable Model**
Image-to-Image	✅	✅
Video-to-Video	✅	❌
Arbitrarily Face Swap	❌	✅
Can Handle Occlusions	❌	✅
Blending Approach	Modified Multi-band blending (Superior)	Poisson blending (Inferior)

Questions?

Sources

- Papers With Code. [n. d.]. Laplacian Pyramids. https:// paperswithcode.com/method/laplacian-pyramid [Online; accessed 18-October-2022].
- Bernhard Egger. 2021. SIGGRAPH2021 - 3D Morphable Face Models - Past, Present and Future - Presentation. https://www.youtube.com/ watch?v=UGtIwWv1dds [Online; accessed 1-October-2022].
- C. Schroers J. Narunic, L. Helminger and R.M. Weber. 2020. High-Resolution Neural Face Swapping for Visual effects. *Computer Graphics Forum* 39, 4 (2020), 173–1
- Yuval Nirkin, Iacopo Masi, Anh Tran Tuan, Tal Hassner, and Gerard Medioni. 2018. On Face Segmentation, Face Swapping, and Face Perception. *In 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*. 98–105. https://doi.org/10.1109/ FG.2018.00024
- Richard J. Radke. 2012. Computer Vision for Visual Effects. Cambridge University Press. https://doi.org/10.1017/CBO9781139019682
- Wikipedia. 2022. Image Gradient. https://en.wikipedia.org/wiki/ Image_gradient [Online; accessed 2-October-2022].

Figure Sources

- Slides 2, 16, 17, 21 images: C. Schroers J. Narunic, L. Helminger and R.M. Weber. 2020. High-Resolution Neural Face Swapping for Visual effects. *Computer Graphics Forum* 39, 4 (2020), 173–184
- Slide 3 image: Creator of Tom Cruise deepfakes shares how he made those viral TikTok videos - CNET
- Slide 5 image (Facial Landmarks): Facial Landmark Detection | Papers With Code
- Slides 5, 19 images (Face Segmentation): Yuval Nirkin, Iacopo Masi, Anh Tran Tuan, Tal Hassner, and Gerard Medioni. 2018. On Face Segmentation, Face Swapping, and Face Perception. *In 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*. 98–105. https://doi.org/10.1109/ FG.2018.00024
- Slides 6, 7, 8, 10, 11, 12, 13 images : Richard J. Radke. 2012. Computer Vision for Visual Effects. Cambridge University Press. https://doi.org/10.1017/CBO9781139019682
- Slide 9 image: Laplacian Pyramid Explained | Papers With Code
- Slide 15 image: Autoencoders with Keras, TensorFlow, and Deep Learning - PyImageSearch
- Slide 18 image: (21) SIGGRAPH2021 - 3D Morphable Face Models - Past, Present and Future - Presentation - YouTube