Manipulação de Atributos Faciais

Rafael de S. Toledo

rafaeltol@gmail.com

Universidade Federal de Santa Catarina

Introdução

É impossível encontrar um rosto caracterizado de todas as formas desejáveis. Caracterizar um rosto com múltiplas combinações de atributos pode fornecer uma **ampliação da base de dados**.

Deep Learning já mostrou seu grande poder de síntese de imagens artificias [1, 2, 3]. Uma de suas principais vertentes para esse fim é o **Autocondificador Variacional** (VAE) [4]. Com VAE é possível isolar e manipular características de uma cena de forma **não supervisionada**.

Para esse projeto foi usado o **CelebA**, que é um grande dataset pra essa tipo de atividade, contendo mais de 200k imagens públicas de celebridades com rótulos de 40 atributos faciais como presença de óculos, chápeu, sorriso, juventudade, gênero, cabelo loiro, e mais.

Solução Proposta

A solução adotada é um VAE com uma estrutura simétrica entre o codificador e o decodificador.

Há dois decodificadores: um gera a máscara do rosto mais cabelo e chapéu (caso haja), o outro reconstrói a imagem inteira.

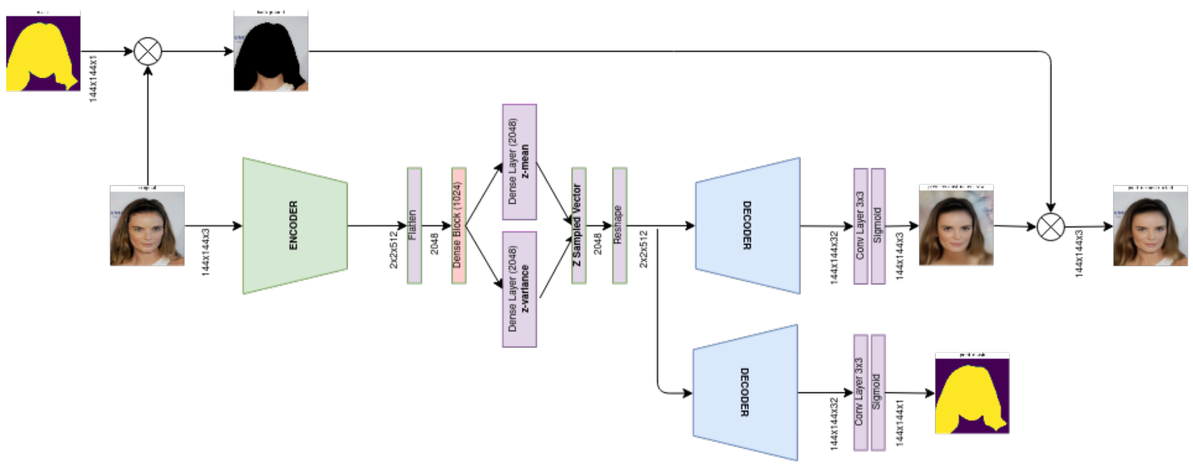

Figure 1: Arquitetura proposta.

Durante o modo de treinamento utiliza-se o rótulo da máscara para substituir os pixels de plano de fundo a fim que as funções custos atuem somente sobre os pixels da face. No modo de predição, a máscara usada será a gerada pelo própria rede.

Há 145 camadas na rede neural, sendo essas convolucionais, *batch normalization*, ativação, e densas. Somando 16.6M de parâmetros.

As funções perdas são divididas em três partes:

1. Reconstrução da imagem: SSIM (*Structural Similarity Index Measure*) e MAE (*Mean Absolute Error*).
2. Reconstrução da máscara: *Binary Cross Entropy* e *DICE loss*.
3. Regularização do espaço latente Z: divergência de Kullback-leibler. Essa escalada no fator de 1/1000.

Treinamento

O treinamento teve 113 steps de 400 iterações de *batch size* 32. Durante mais de 50 horas.

Usou-se o otimizador Adam e taxa de aprendizado de 1e-4, cortando gradiente no máximo de 1e-3, e learning decay de 1/3 a cada 10 épocas até atingir o mínimo de 1e-5.

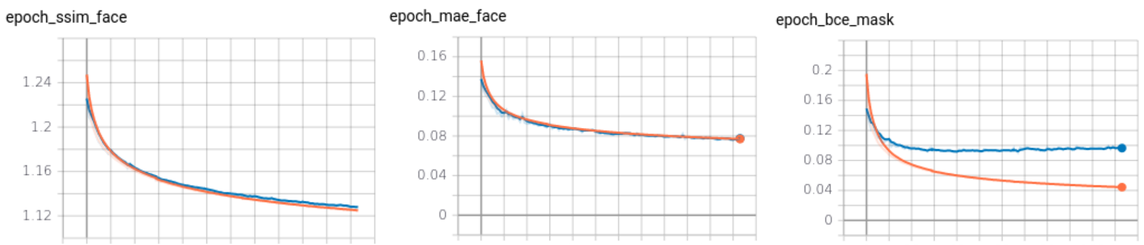

Figure 2: Performance das funções custos ao longo do treinamento, seguindo respectivamente: $2 - SSIM$, MAE e o $BCE + Diceloss$ sobre a máscara.

Aritmética no Espaço Latente

Em [2] mostrou-se que os atributos podem ser representados por aritmética no espaço latente.

Para cada característica foi pego uma amostragem de 10k imagens de exemplos positivos, e 10k de exemples negativos. Computou-se a média de ambos vetores, e no final a diferença das médias.

Após isso para edição do atributo, adiciona-se o vetor do atributo aos valores correntes do espaço latente da amostra sendo predita.

Conclusões

- Para a reconstrução da imagem, constatou-se pouquíssima presença de *variância* no dataset celebA, não tendo diferença de desempenho entre resultados no conjunto de treinamento ou de teste.
- Mostrou-se possível a adição de atributos específico dentro de uma figura sem corromper o resto da imagem original.
- A inserção ou alteração de atributos faciais tendem a se alinhar a pose a expressão do rosto original da foto.
- Detalhes sutis tendem a ficar borrados na reconstrução da imagem. Sendo um ponto de melhora para trabalhos futuros.

Espectro do Atributo

É possível manipular o atributo dentro de uma escala contínua de intensidade em um espectro posivito e negativo. Resultando em um espectro de combinações do rosto com aquele atributo.

Figure 3: Primeira coluna é a image original. Cada linha representa um atributo, seguindo na ordem: Sorriso, Juventude, Loira, Óculos.

Edição de Atributos

Para uma manipulação mais direta, segue-se adição de atributos específicos a seus exemplos negativos.

Na figura 4 são ilustrados 4 processos de criação de barba. Na segunda coluna usou-se a escala negativa do atributo genérico de *No_beard*. As colunas em seguida com colaração foi computado o vetor atributo especificamente de subgrupos de pessoas com barba grisalha, preta e loira.

Figure 4: Adicionando barbas de diferentes colarações.

Figure 5: Adicionando os respectivos atributos ao rosto feminino: estreitar olhos, adicionando óculos, cabelos loiros, e adicionando maquiagem.

Referências

[1] Tero Karras, Samuli Laine, and Timo Aila. “A style-based generator architecture for generative adversarial networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 4401–4410.

[2] Alec Radford, Luke Metz, and Soumith Chintala. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2016. arXiv: 1511.06434 [cs.LG].

[3] Ting-Chun Wang et al. “Video-to-video synthesis”. In: *arXiv preprint arXiv:1808.06601* (2018).

[4] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013).