Problem:
Generative Adversarial Network (GANs) has been extensively used in machine learning community to create new faces, scenes, and applying style transfers. In this project, we modified two GANs based architecture to generate realistic face sketches, given an input domain of cropped photos.

Contributions:
- Introduced edge loss term to reduce “adding glasses” problem.
- Compiled a dataset for unsupervised training of face sketch generation.

Unsupervised Generative Attentional Networks

Major differences from a conventional GAN:
- Attention module (auxiliary classifier): $\eta_{a}(x) = \sigma(\sum_{k} w_{k}^{a} \sum_{i,j} E_{a}^{i,j}(x))$
- Adaptive Layer-Instance Normalization:
  $a_{i} = \frac{a - \mu_{i}}{\sigma_{i}^{2} + \epsilon}$
  $a_{k}^{l} = \frac{a - \mu_{k}^{l}}{\sigma_{k}^{2} + \epsilon}$
  $AdaLIN(a, \gamma, \beta) = \gamma \cdot (\rho \cdot a_{i} + (1 - \rho) \cdot a_{k}^{l}) + \beta$
  $\rho \leftarrow clip(\rho, (\rho - \tau \Delta \rho))$

Loss function

$$\min_{G_{s} \rightarrow G_{t}, G_{t} \rightarrow G_{s}, \eta_{s}, \eta_{t}, D_{s}, D_{t}, \theta_{D_{s}}, \theta_{D_{t}}} \max_{D_{s}, D_{t}} \lambda_{1} L_{gan} + \lambda_{2} L_{cycle} + \lambda_{3} L_{identity} + \lambda_{4} L_{cam} + \lambda_{5} L_{edge}$$

Loss details

$L_{cycle} = \mathbb{E}_{x \sim X_{s}}[||x - G_{t \rightarrow s}(G_{s \rightarrow t}(x))||_{1}])$

$L_{identity} = \mathbb{E}_{x \sim X_{s}}[||x - G_{s \rightarrow t}(x)||_{1}]$

$\mathbb{E}_{x \sim X_{s}}[log(\eta_{s}(x))]$$ + \mathbb{E}_{x \sim X_{t}}[log(1 - \eta_{t}(x))]$

$\mathbb{E}_{x \sim X_{s}}[\eta_{s}(D_{s}(x))] + \mathbb{E}_{x \sim X_{t}}[log(1 - \eta_{t}(G_{s \rightarrow t}(x)))]]$

Encourage alignment of edges between a face and sketch

Color distributions of input and output images should be similar

Translates of image $X_{s}$ to $X_{t}$ and back should preserve image

Training and Results

Dataset: For training and testing, we used the Chinese University of Hong Kong’s Face Sketch Database to get images for sketches. For images of faces, we used the Chicago Face Dataset and the IMM Face dataset. For training, we used around 500 images and 150 images for testing.

Experiment: We used the data to train our proposed network and used results trained on CycleGAN as a baseline to compare our results with. We obtain quantitative results using a testing set from CUHK which has faces and corresponding sketches made by an artist.

Quantitative results:

<table>
<thead>
<tr>
<th></th>
<th>CycleGAN</th>
<th>U GAT IT (proposed loss function)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural Similarity (SSIM)</td>
<td>0.556</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Qualitative results:

U-GAT-IT prediction

CycleGAN prediction

Conclusion: U GAT IT performs much better than CycleGAN. It does not add unnecessary artefacts, and preserves the shape and expression of the face. The higher SSIM backs up the claim.

References

