Visual Computing

Masked Unmasked Face Recognition

Group 9

2021-07-28

Outline

- Introduction of group members
- Levels we achieved
- Our Work in Bonus Level
- Future works
- Q & A

Introduction of Group Members

- Chen Jinghao （陈景浩）: Sophomore, SJTU, Information Engineering
- Liu Yongxiang （刘永翔）: Sophomore, SJTU, Computer Science
- Huang Zhemin （黄喆敏）: Sophomore, SJTU, Software Engineering
- Gao He （高贺）: Junior, HZAU, Automation

Division of labour

- Chen Jinghao （陈景浩）： Bonus level, beginner & expert level
- Liu Yongxiang （刘永翔）： Beginner & expert level, slides
- Huang Zhemin （黄喆敏）： Presentation, beginner & expert level
- Gao He （高贺）： Poster, beginner level

Levels we achieved

- In our project, we achieved the **bonus level** of the work.
- In general, we got a **HIGH** accuracy of face recognition.
- For beginner & expert level, please refer to the poster.

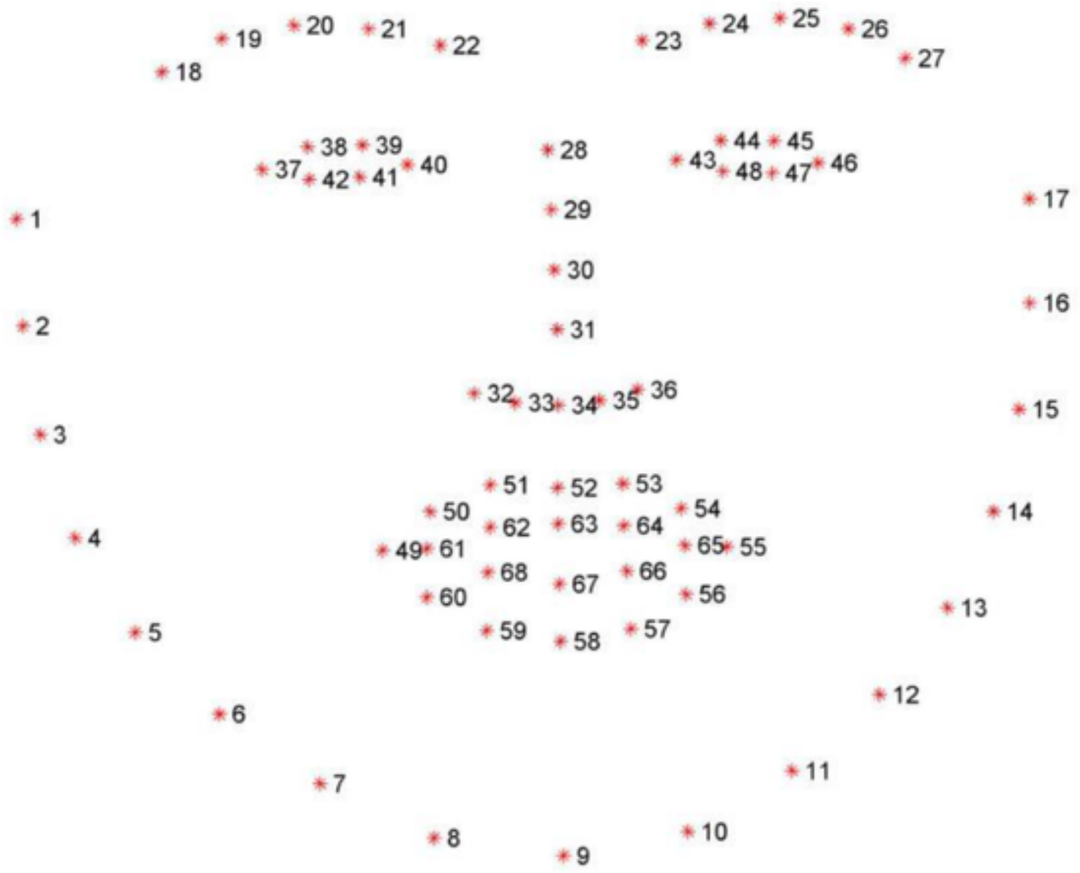
Our Work in Bonus Level

Preprocessing

For Georgia Tech Face database,

- Use dlib to detect the face.
- Crop the face and resize it to 150*150 pixels.
- Keep 3 channels (RGB), use gaussian blur to reduce noise.
- Use 'shaped predictor 68 face landmarks' to grab key points.
- Split datasets into training sets (80%) and testing sets (20%).

Preprocessing



Preprocessing

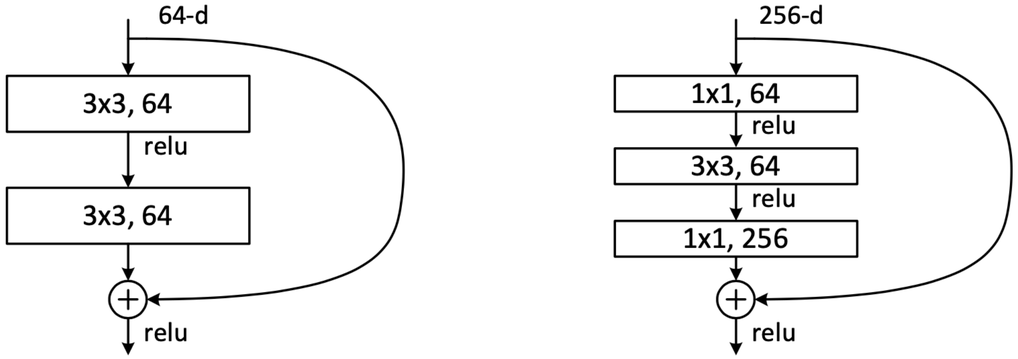
For celebA face dataset,

- Almost the same.
- But only choose label range from 0 to 100.
- The size of processed dataset is about 20*100.

Features

For both datasets, use deep residual learning to extract features. PyTorch

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9



A deeper residual function for ImageNet. Left: A original building block. Right: a “bottleneck” building block for ResNet-50/101/152.

[1] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Features

For Georgia Tech Face database,

- Backbone: resnet18
- Optimizer: Adam
- Loss function: cross-entropy loss function
- Since there are 50 subjects in the dataset, we set the dimension of the last fully connected layers as 50.

Features

For celebA face dataset,

- Backbone: resnet-18 vs ~~wide_resnet101~~
- Dimension of last fully-connected layer: 100 ~~50~~ (100 subjects)
- Strategy of reducing learning rate: ReduceLROnPlateau

Classifiers

- SVM.
- Use resnet18 & wide_resnet101 to train the classifier by unmasked dataset features.
- Prediction & Comparison.

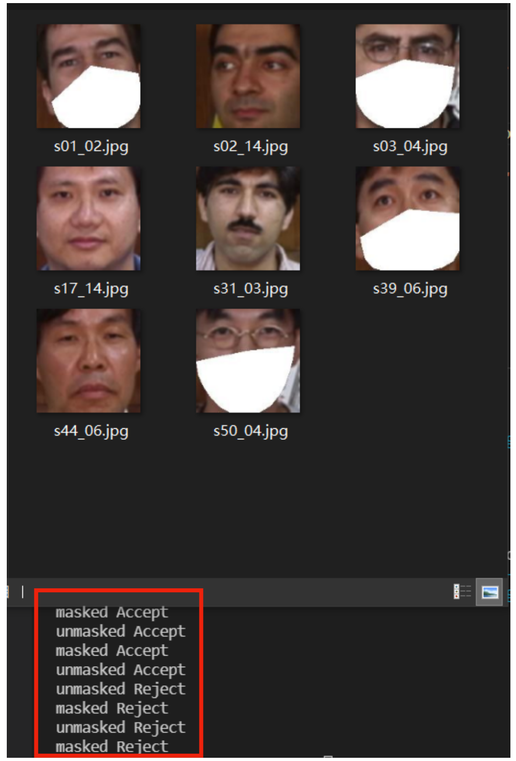
Results

- For Georgia Tech Face database, we achieve **nearly 100%** accuracy. Only 1 out of 750 was wrongly predicted.
- For celebA face dataset, We achieve **73.3%** accuracy in the masked dataset and **75.2%** accuracy in the unmasked dataset.
- The accuracy of judging mask is still **100%**.
- Considering CIFAR-100 only achieved about 75% accuracy[1], our result is an ideal one.

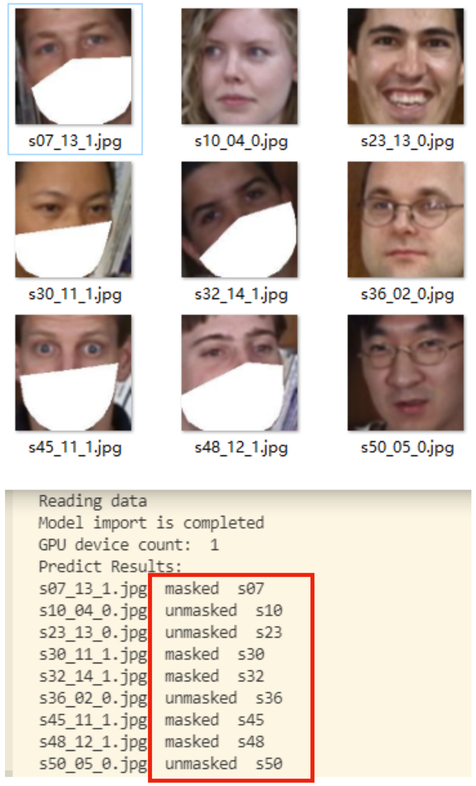
[1] Reference: http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html#43494641522d313030

Results

- Examples of successful classifications



Expert Level



Bonus Level

Result of Georgia Tech Face database

Results

- Examples of successful classifications

```
Model import is completed
GPU device count:  3
Predicted_Results: masked  13
Predicted_Results: masked  35
Predicted_Results: unmasked  27
Predicted_Results: unmasked  55
Predicted_Results: unmasked  46
Predicted_Results: unmasked  66
Predicted_Results: masked  4
Predicted_Results: masked  27
Predicted_Results: masked  33
Predicted_Results: unmasked  31
Predicted_Results: unmasked  41
(yolo) danielchen@amax:~/nus$
[nus] 0:bash*
```


000077_13.jpg


001265_35.jpg


003775_27.jpg


004198_57.jpg


004234_46.jpg


004267_66.jpg


022543_4.jpg


022588_27.jpg


022710_33.jpg


160527_31.jpg


160665_41.jpg

Result of celebA face dataset

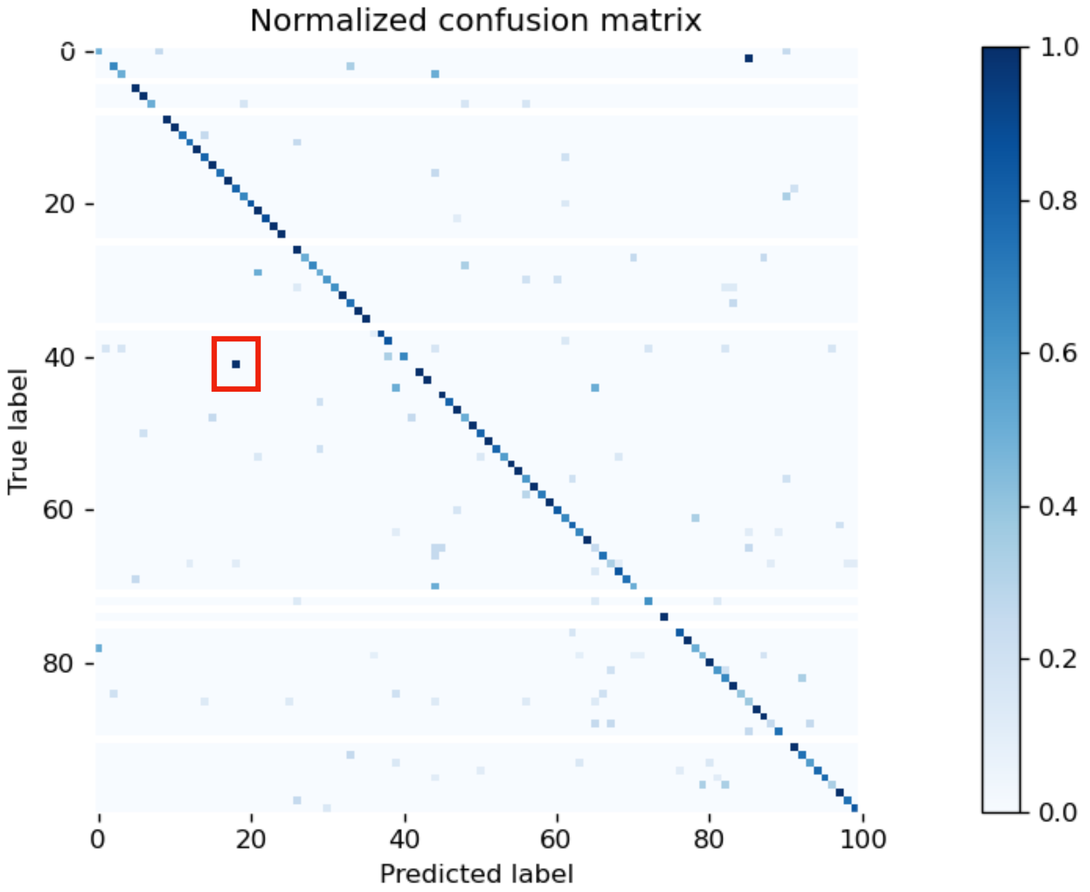
Future works

- There remain some unsuccessful and challenging classifications.
- e.g. ‘Twins’ after wearing masks, similar types of lighting, similar facial expressions, etc.

s45

s07

Future works



s41

s18

Thanks for Watching

Group 9

Q&A