Facial Landmarks Detection with Fake-it Dataset

CMPT 732 Project

–Jingwen Che, Haoran Ding, Chen Qiao

Overview

- Design a deep learning neural network model for facial landmarks detection
- Train the model on the CG faces
- Test the model on the real world faces

Our output

Real world samples of human faces with facial landmarks annotated.

Best Model

Our Model !

MSE Loss	Training loss	Validation Loss	Test Loss
Our Model	0.00078196	0.00007003	0.00269198
Xception	0.00083827	0.00002770	0.00287737
ResNet-50	0.00187827	0.00046778	0.00385637
MobileNet-v2	0.00116653	0.00011362	0.00336023

Dataset

- Training data:

  CG dataset from Microsoft

  (with 70 standard facial landmarks)

Dataset

- Testing data:

  Flickr-Faces-HQ Dataset

  Video (30 fps)

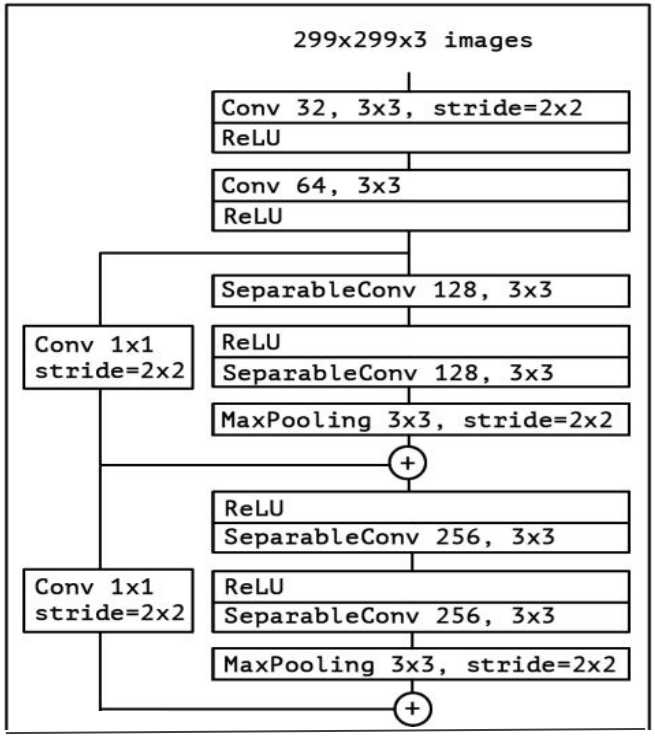
Data preprocessing

- Box the Face
- Original size: 512*512
- After resized: 128*128
- Implemented through

  TF.resize and TF.crop

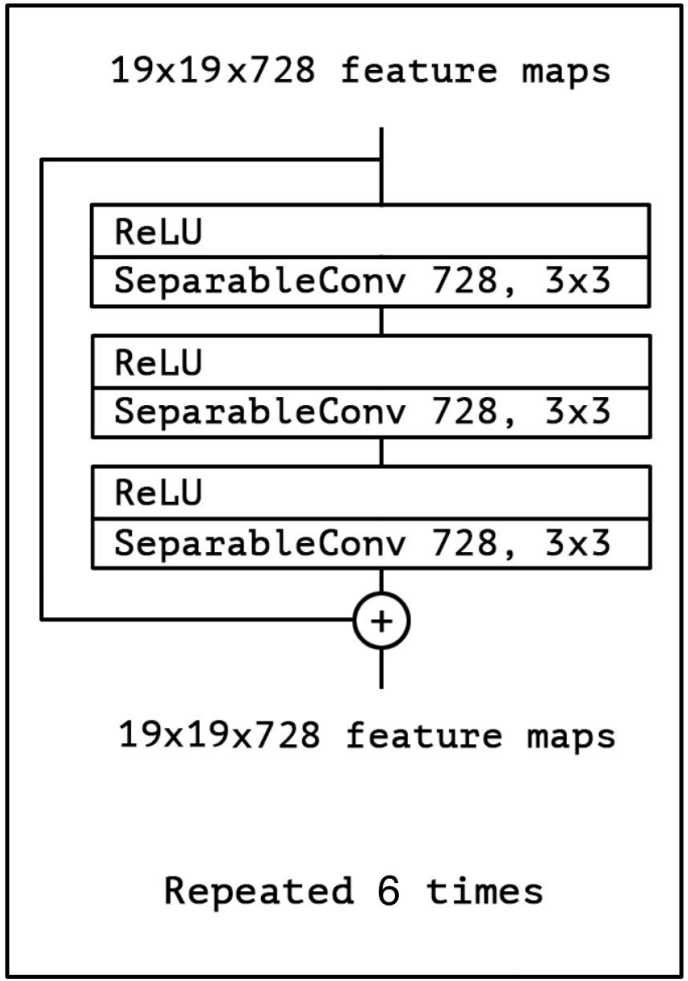
Data augmentation

- Random Gamma
- Random Contrast
- Random Hue
- Random Saturation
- Random Brightness
- Random Rotation



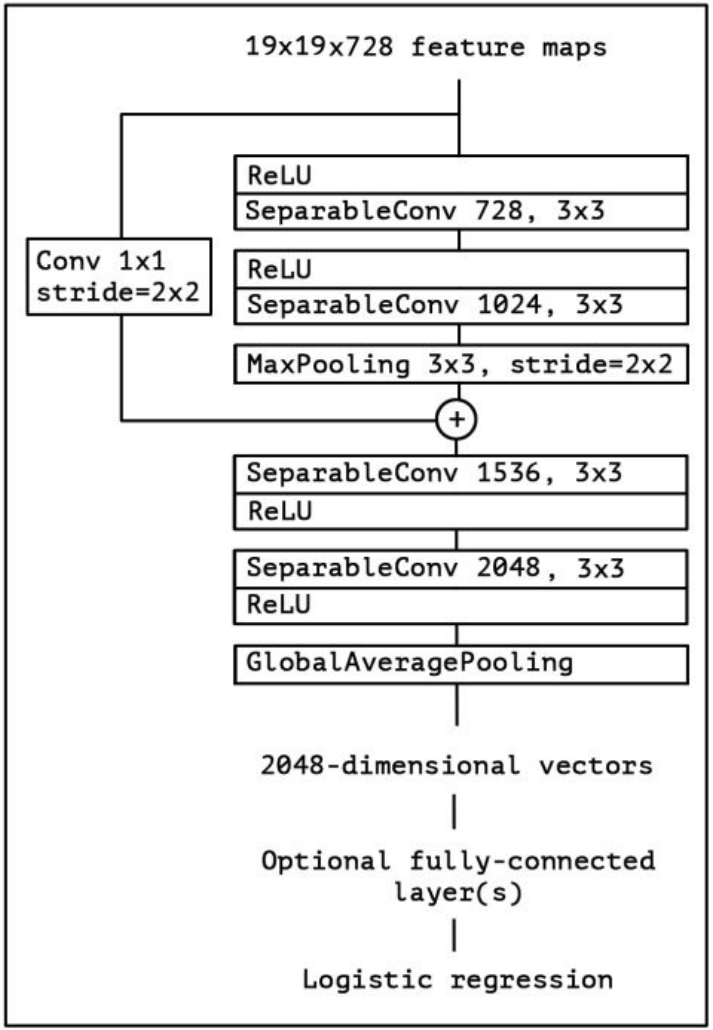
Network architectures

- **Our Model**
  - **Entry Flow (two residual)**
  - Middle Flow
  - Exit Flow



Network architectures

- **Our Model**
  - Entry Flow
  - **Middle Flow**
  - Exit Flow

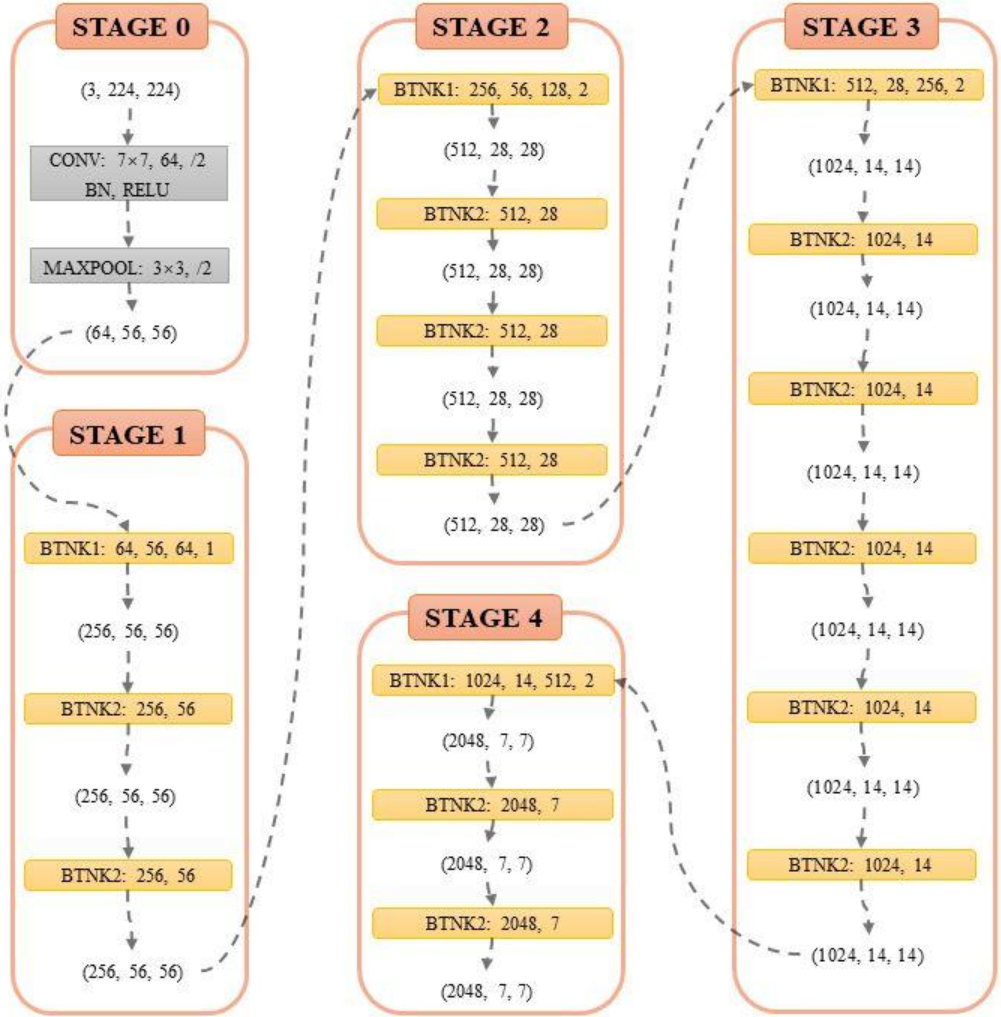


Network architectures

- **Our Model**
  - Entry Flow
  - Middle Flow
  - **Exit Flow**

Network architectures

- **ResNet-50**
  - Stage 0
  - Stage 1
  - Stage 2
  - Stage 3
  - Stage 4



Network architectures

- **MobileNet_v2**
  - Inverted residual
  - Bottleneck residual block

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Result

- Our model

  Validation set prediction

Result

- Our model

  Test set prediction

Result

- Different Versions of Xception Network (5000 images in training set)

MSE Loss	Training loss	Validation Loss	Test Loss
Our Model LeakyReLU & 6 mid	**0.00174931**	**0.00010353**	**0.00334268**
Xception	0.00155704	0.00010563	0.00339070
Xception with LeakyReLU	0.00149869	0.00018571	0.00349785
Xception with 6 mid block	0.00386720	0.00052409	0.00654623

Result

- Results of Different training set sizes
- 5000 is a viable size

Validation MSE Loss	Training set 2000	Training set 5000	Training set 10000
Our Model	**0.00070798**	**0.00010353**	**0.00003220**
Xception	0.00091584	0.00010563	0.00003259
ResNet50	0.00120577	0.00018337	0.00007236
MobileNet-v2	0.00045519	0.00023415	0.00006439

Result

- Results of our model, Xception network, ResNet-50, and MobileNet_v2 after Hyperparameter optimization

MSE Loss	Training loss	Validation Loss	Test Loss
Our Model	**0.00078196**	**0.00007003**	**0.00269198**
Xception	0.00083827	0.00002770	0.00287737
ResNet50	0.00187827	0.00046778	0.00385637
MobileNet-v2	0.00116653	0.00011362	0.00336023

What we will do for the remaining time?

- Refine the code to make it clean and robust
- Try to implement data mixing during training to improves the result
- Pictorially demonstrate the results and complete the report

Reference

- Erroll Wood, Tadas Baltrusaitis, Charlie Hewitt, Sebastian Dziadzio, Thomas J. Cashman, Jamie Shotton.: Fake it till you make it: face analysis in the wild using synthetic data alone. International Conference on Computer Vision 2021. https://openaccess.thecvf.com/content/ICCV2021/html/Wood_Fake_It_Till_You_Make_It_Face_Analysis_in_the_ICCV_2021_paper.html

- ChihFan Hsu, ChiaChing Lin, TingYang Hung, ChinLaung Lei, KuanTa Chen: Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization. arXiv:2005.08649. https://arxiv.org/abs/2005.08649

Thank you!