جامعة خليفة
Khalifa University

Facial Expression Recognition with Attention Technique

Adarsh Ghimire (100058927)
Selina Shrestha (100058926)
Youssef Ibrahim (100058639)

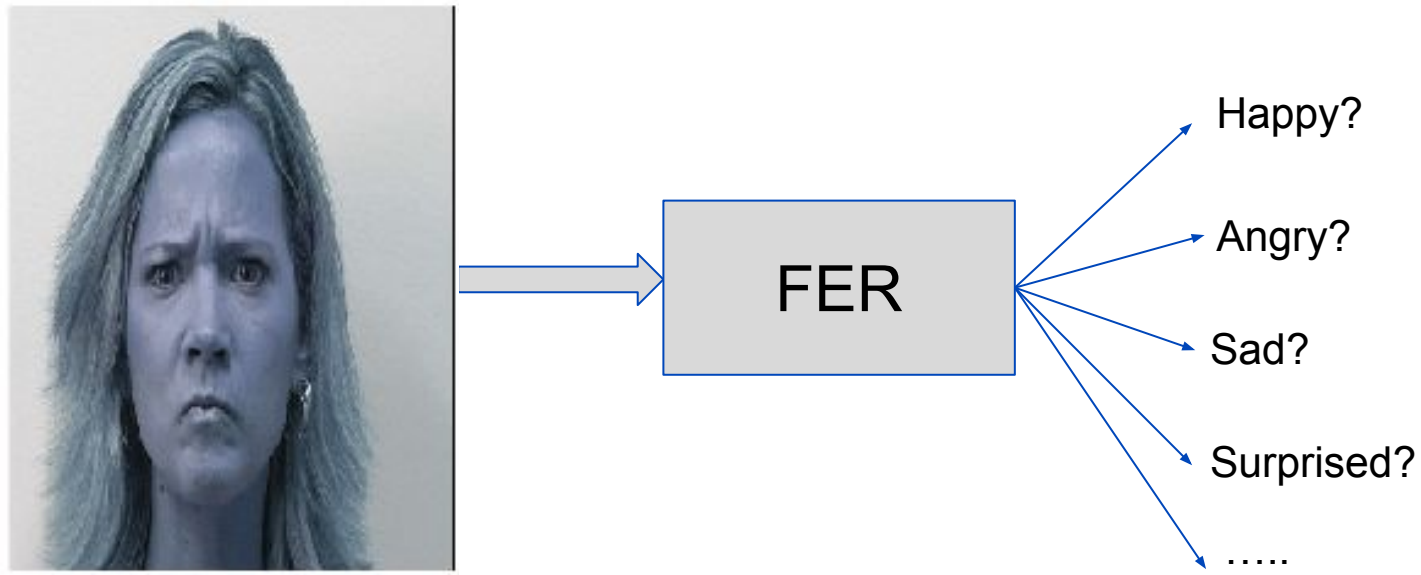
30 NOVEMBER 2020

Outline

- Introduction
- Literature Review
- Research Questions
- Methodology
- Results & Discussion
- Conclusion and Future works

Introduction

What is Facial Expression Recognition (FER)?



source: [8]

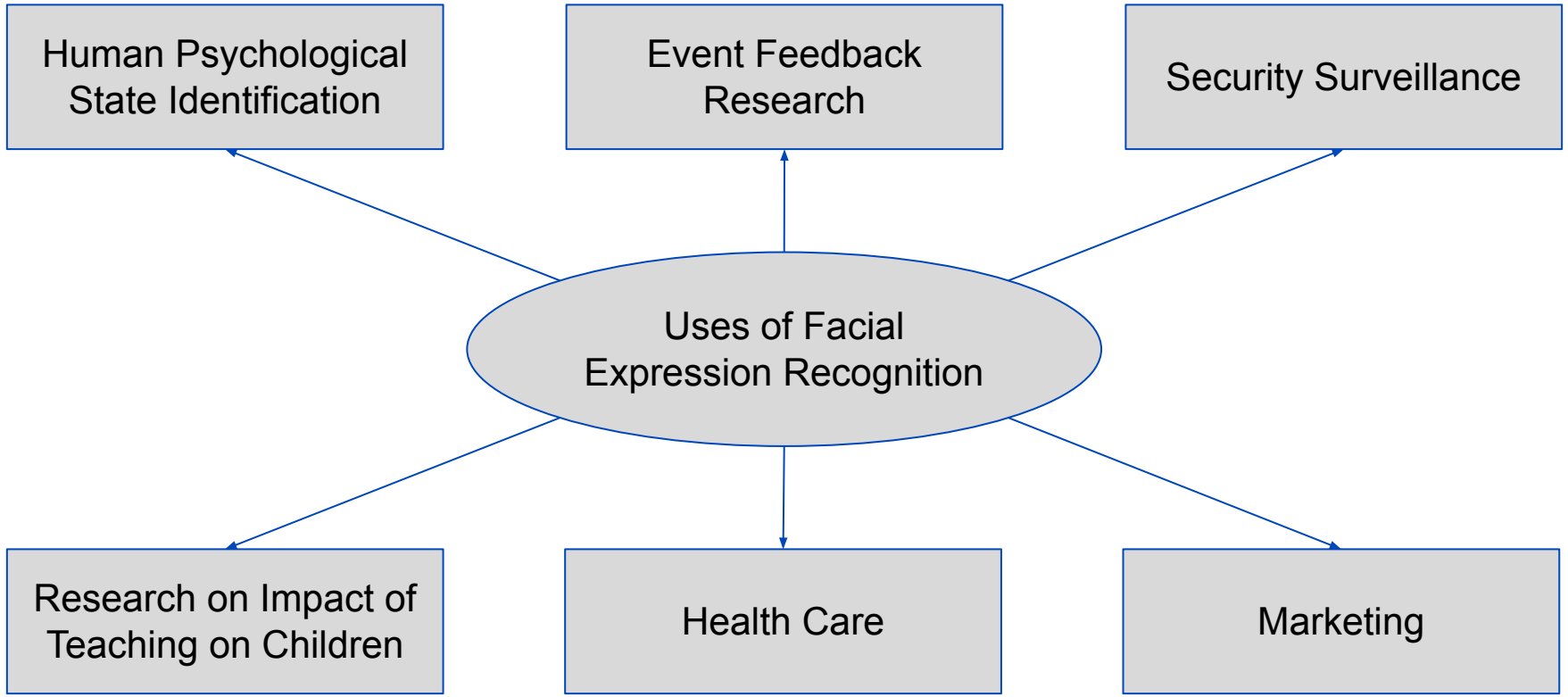
Introduction

How is Facial Expression Recognition different than other Image classification tasks?

- Certain areas of face are used more while expressing feelings
- So, certain parts of face play bigger role in identifying expressions
- Need for mechanism to identify important regions and assign greater weights to such features

source: [8]

Introduction



Literature Review

Earlier Works: Handcrafted features and traditional Machine Learning

Paper	Features	Classifier
[1]	Histogram of oriented gradients (HOG)	SVM
[2]	Local binary patterns (LBP), Haar Features	Random Forest
[3]	Local binary patterns (LBP), Gabor wavelets	SVM

Later Works: Deep Neural Networks and Convolution Neural Networks

Paper	Model Description
[4]	CNN with residual blocks
[5]	Multiple CNN networks trained in parallel and ensembled
[6]	2 parallel feature extraction blocks inspired by GoogleNet

Recent Works: Convolution Neural Networks with Attention Network

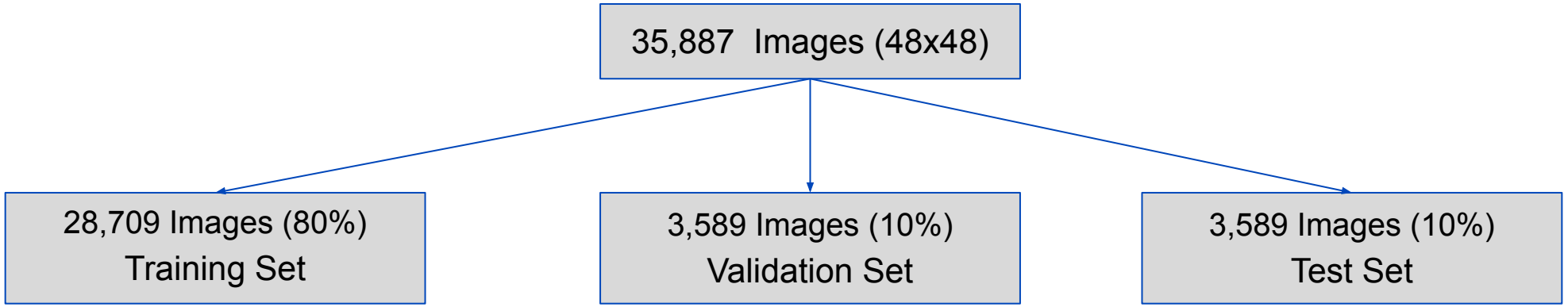
Paper	Model Description
[7]	Images are cropped and passed to Region Attention Network
[8]	Attentional Convolutional Network using Spatial Transformer Network

Research Questions

- **Which CNN architecture works best for Facial Expression Recognition task?**
- **Can the addition of attention network improve the accuracy on facial expression recognition?**
- **Where can we add the attention network?**

Dataset

- **Facial Expression Recognition 2013 (FER-2013) Dataset**
  - contains partial and occluded faces
  - previous works have not been able to achieve very high accuracies

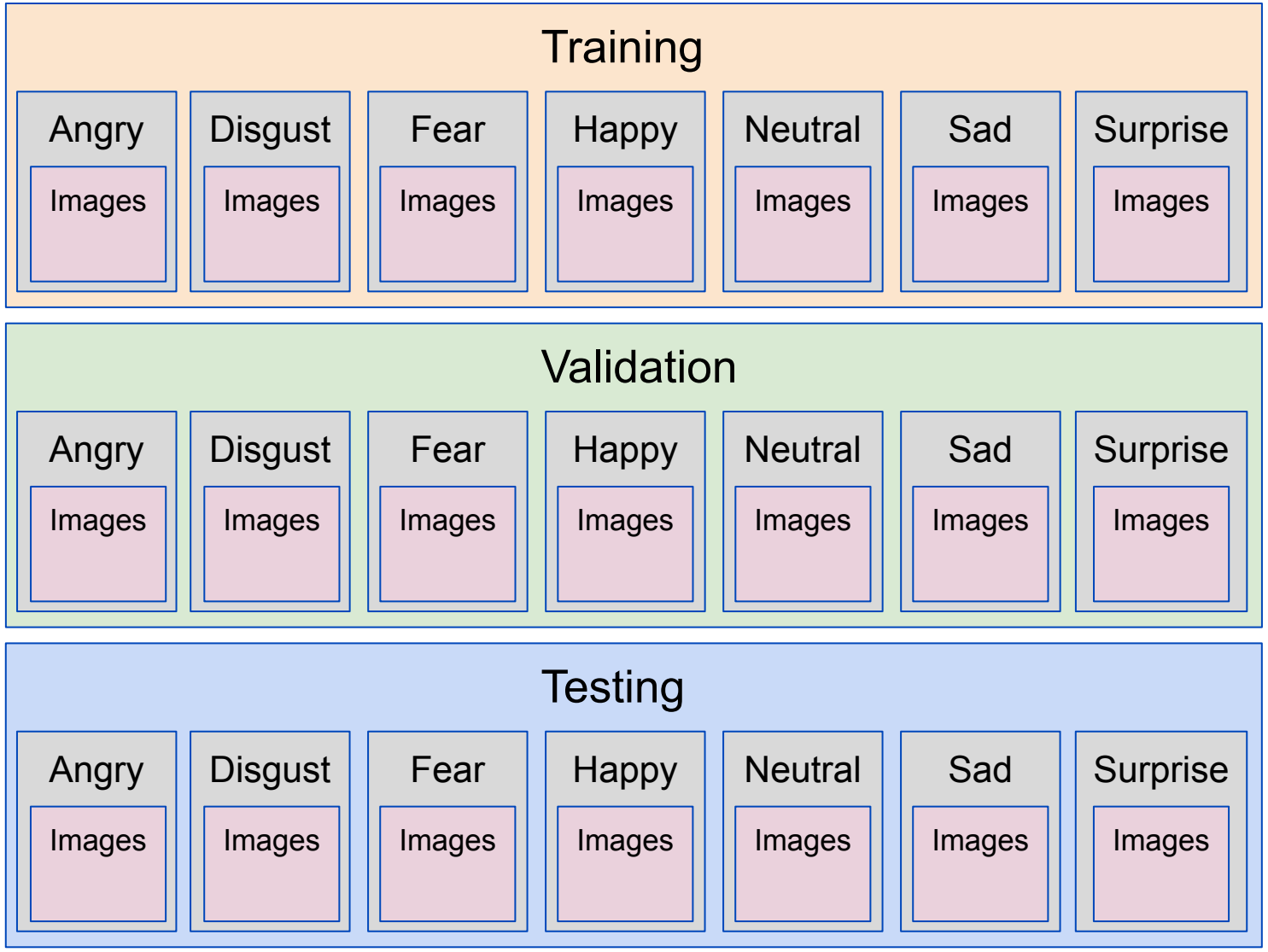


Set/ Classes	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Training	3,995	436	4,097	7,215	4,965	4,830	3,171
Validation	467	56	496	895	607	653	415
Test	491	55	528	879	626	594	416

Methodology

1. Dataset Preparation

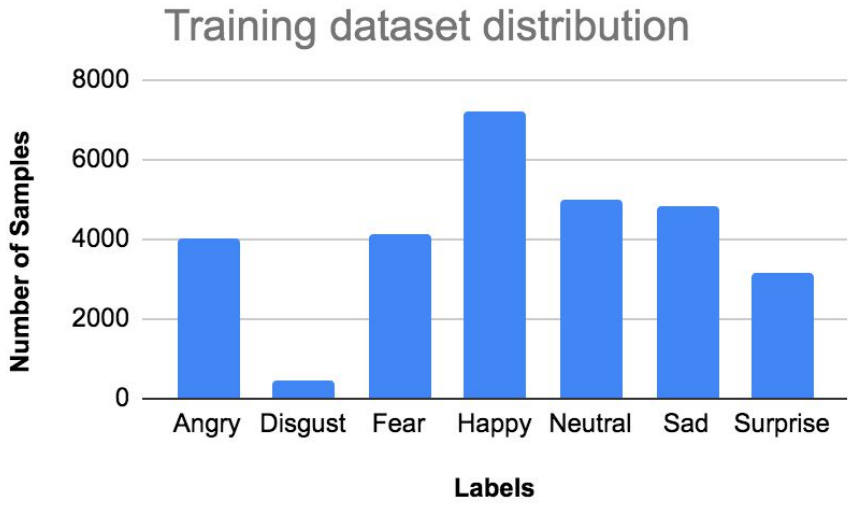
	A	B	C
1	emotion	pixels	Usage
2	0	70 80 82 72 !	Training
3	0	151 150 147	Training
4	2	231 212 156	Training
5	4	24 32 36 30 !	Training
6	6	4 0 0 0 0 0 0	Training
7	2	55 55 55 55 !	Training
8	4	20 17 19 21 !	Training
9	3	77 78 79 79 !	Training
10	3	85 84 90 121	Training
11	2	255 254 255	Training
12	0	30 24 21 23 !	Training
13	6	39 75 78 58 !	Training
14	6	219 213 206	Training
15	6	148 144 130	Training
16	3	4 2 13 41 56	Training
17	5	107 107 109	Training
18	3	14 14 18 28 !	Training
19	2	255 255 255	Training
20	6	134 124 167	Training
21	4	219 192 179	Training
22	4	1 1 1 1 1 1 1	Training
23	2	174 51 37 37	Training
24	0	123 125 124	Training
25	0	8 9 14 21 26	Training
26	3	252 250 246	Training
27	3	224 227 219	Training



Methodology

2. Data Augmentation and Class weights

Data Augmentation Metrics
Rescaling
Rotation
Width Shift
Height Shift
Horizontal Flip



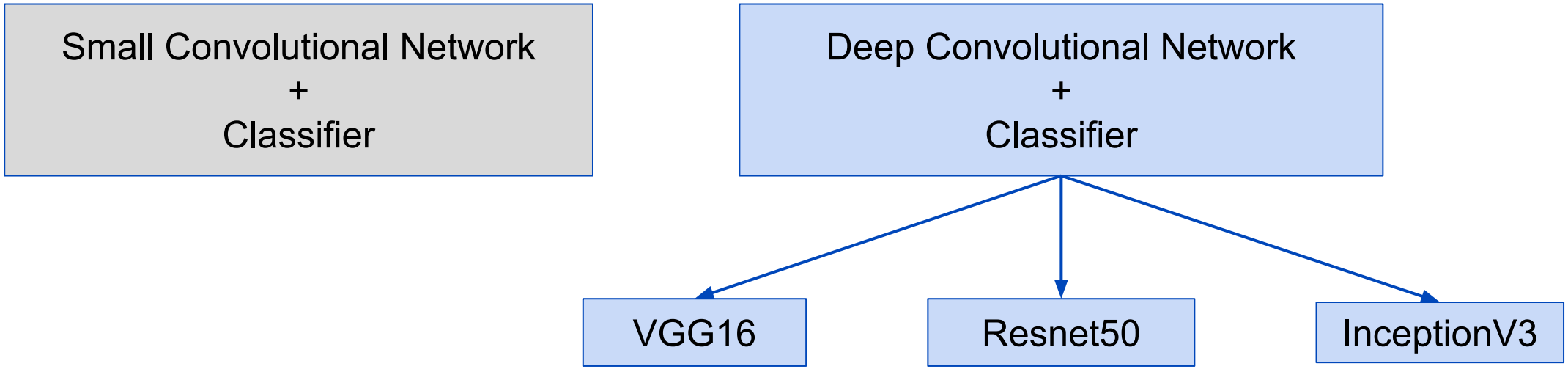
$$category weight = \frac{Total number of training samples}{Number of classes \times Number of samples in the category}$$

Methodology

Major Task

1. Finding the best architecture for transfer learning
2. Adding attention network on top of best architecture

1. Finding the best architecture for transfer learning

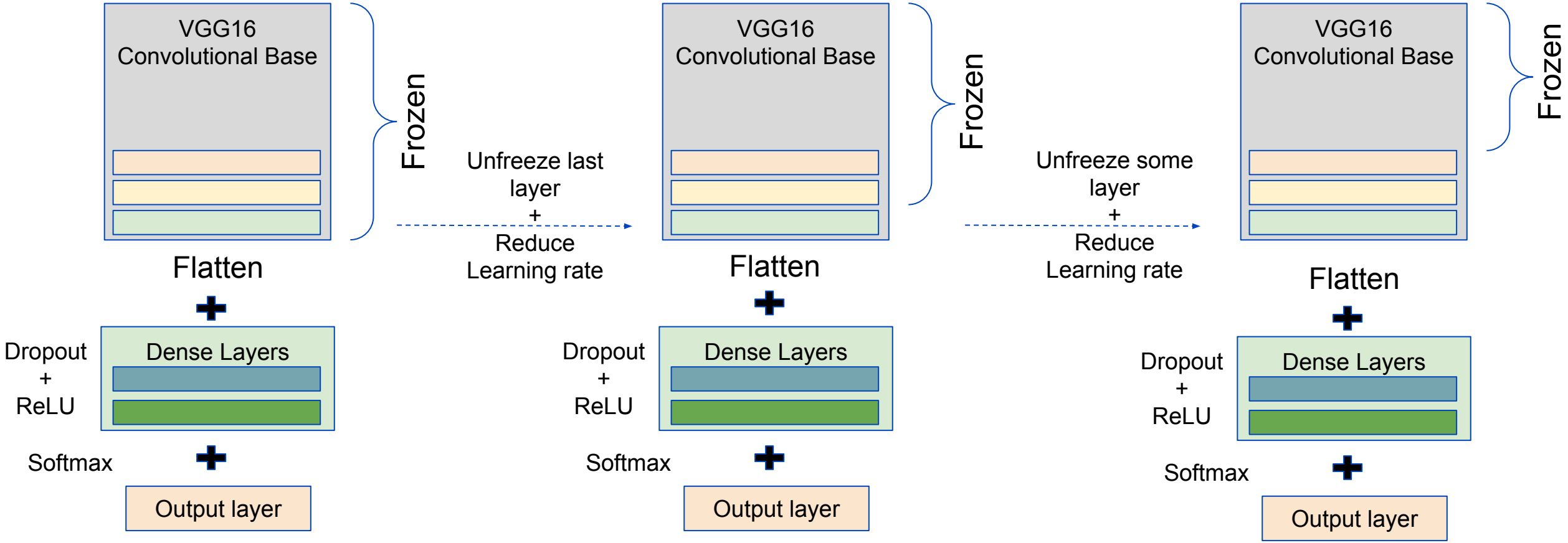


Methodology

1. Finding the best architecture for transfer learning

VGG16

1. VGG16 network pre-trained on Imagenet/ VGG Face

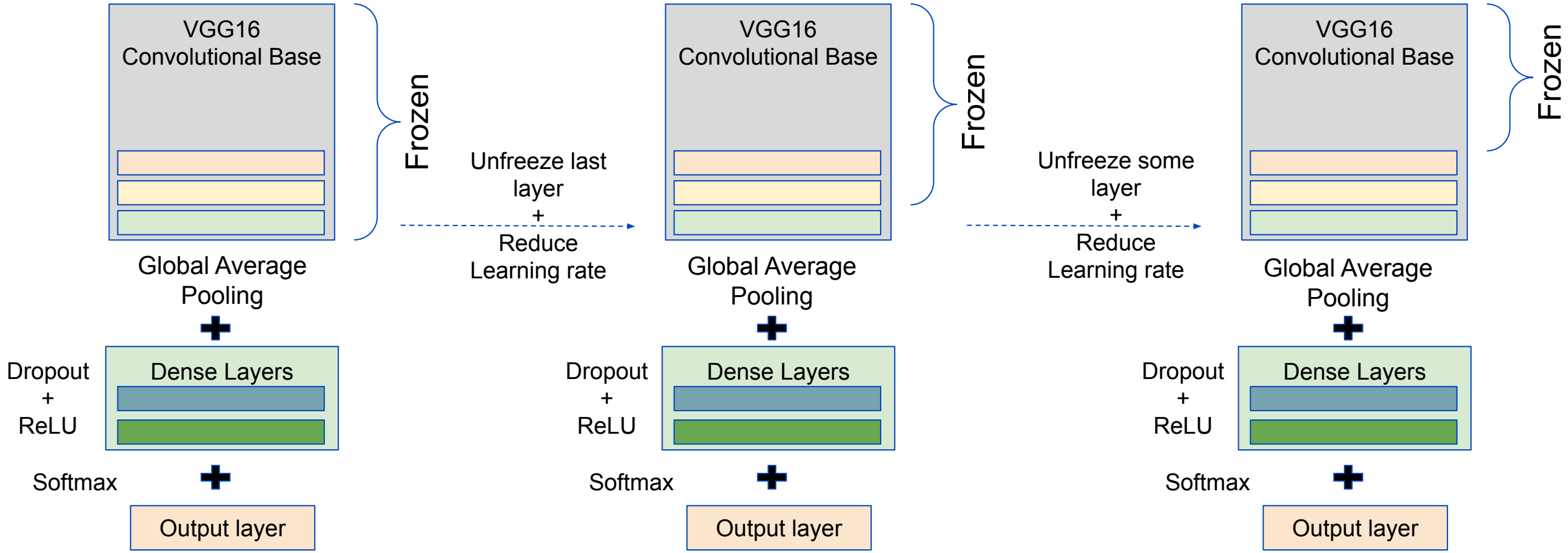


Methodology

1. Finding the best architecture for transfer learning

VGG16

2. VGG16 network pre-trained on VGG Face

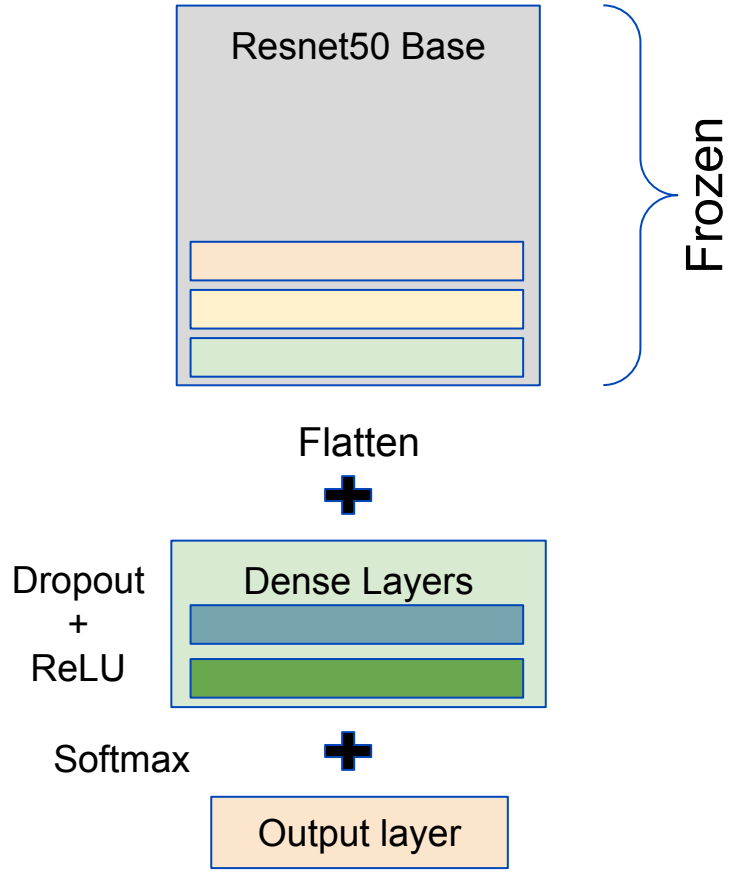


Methodology

1. Finding the best architecture for transfer learning

Resnet50

1. Resnet50 network pre-trained on Imagenet/VGG Face

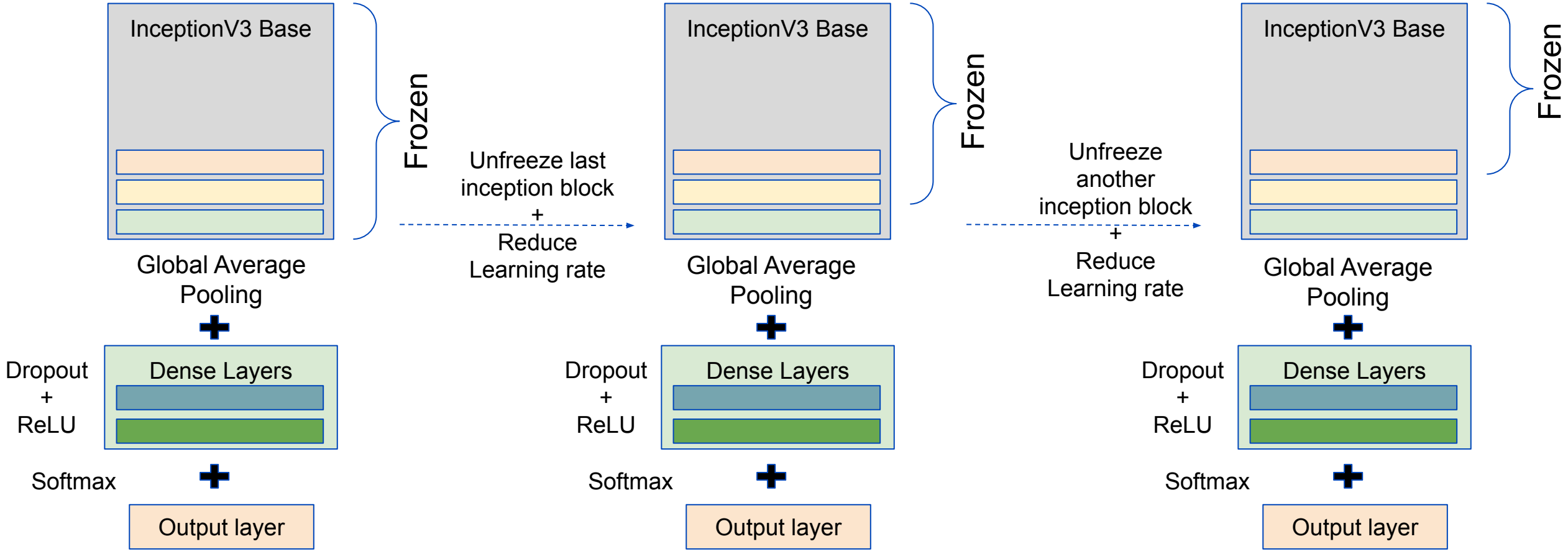


Methodology

1. Finding the best architecture for transfer learning

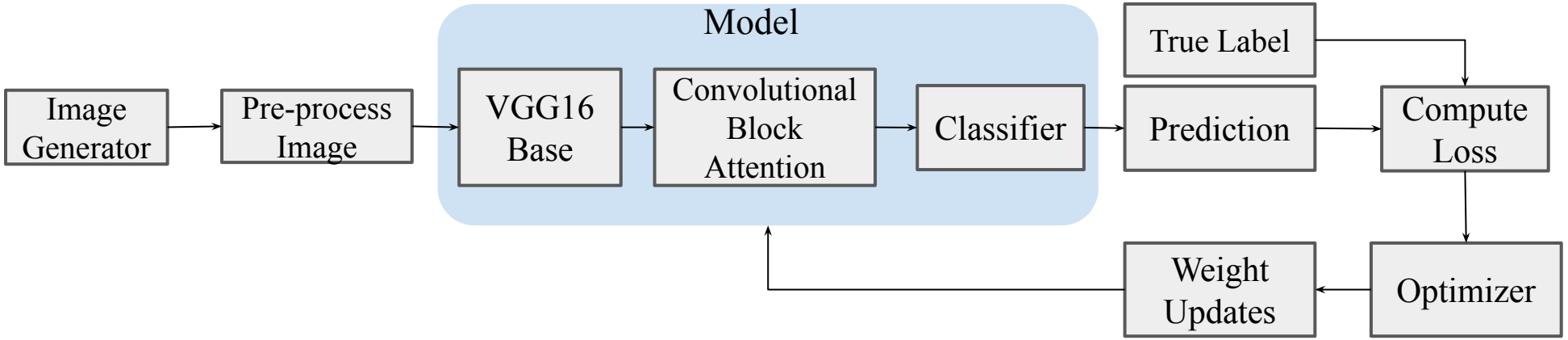
InceptionV3

1. InceptionV3 network pre-trained on Imagenet



Methodology

Adding Attention on top of best performing architecture

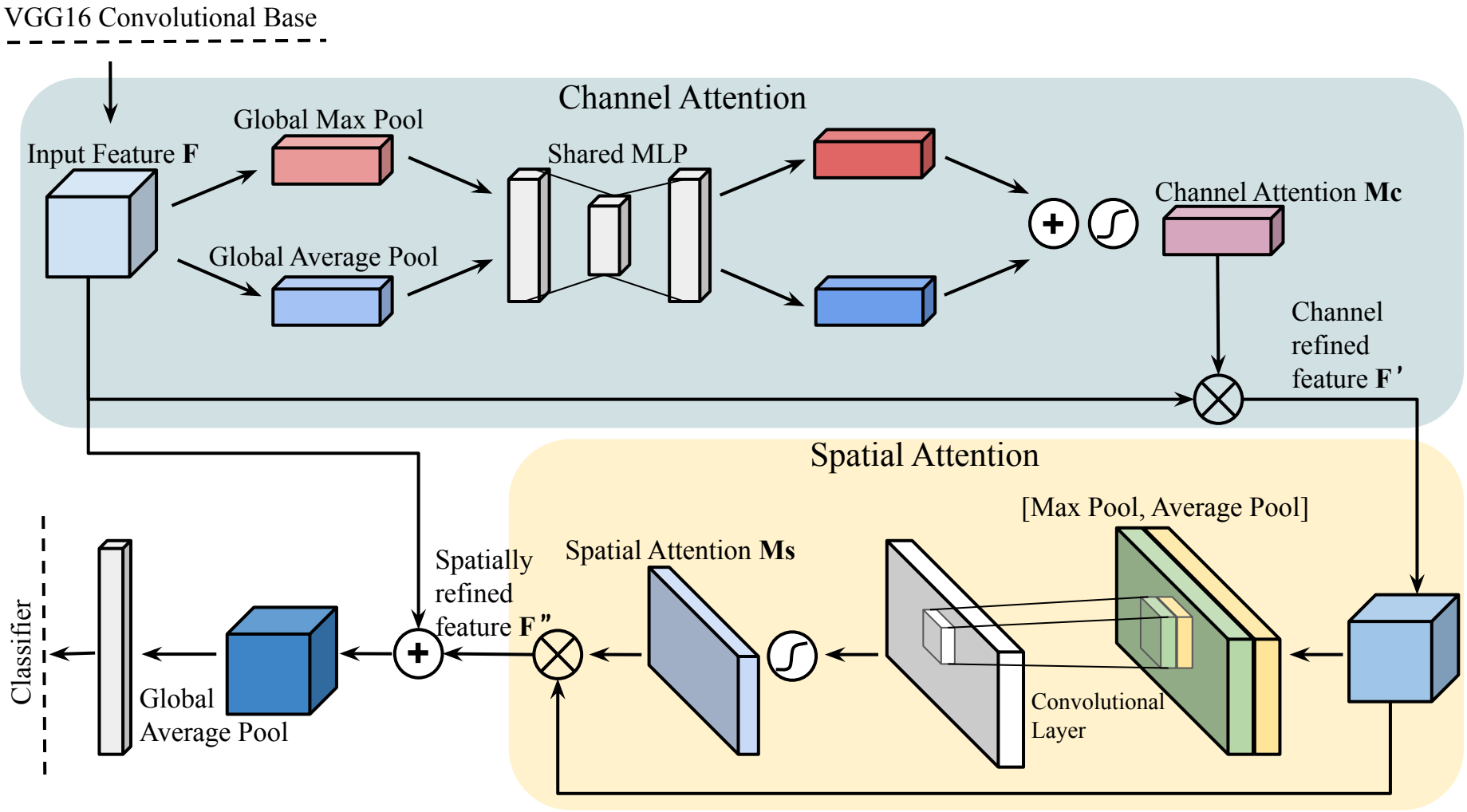


Source : [9]

Input Layer	Input	(?, 224, 224, 3)
	Output	(?, 224, 224, 3)
VGG16 Conv Base Layer	Input	(?, 224, 224, 3)
	Output	(?, 7, 7, 512)
Channel Attention Layer	Input	(?, 7, 7, 512)
	Output	(?, 1, 1, 512)
Multiplying Layer	Input	[(?, 1, 1, 512), (?, 7, 7, 512)]
	Output	(?, 7, 7, 512)
Spatial Attention Layer	Input	(?, 7, 7, 512)
	Output	(?, 7, 7, 1)
Multiplying Layer	Input	[(?, 7, 7, 1), (?, 7, 7, 512)]
	Output	(?, 7, 7, 512)
Addition Layer	Input	[(?, 7, 7, 512), (?, 7, 7, 512)]
	Output	(?, 7, 7, 512)
Global Average Pooling Layer	Input	(?, 7, 7, 512)
	Output	(?, 512)
Dense Layer (1) (ReLU)	Input	(?, 512)
	Output	(?, 512)
Dense Layer (2) (ReLU)	Input	(?, 512)
	Output	(?, 512)
Classifier (Softmax)	Input	(?, 512)
	Output	(?, 7)

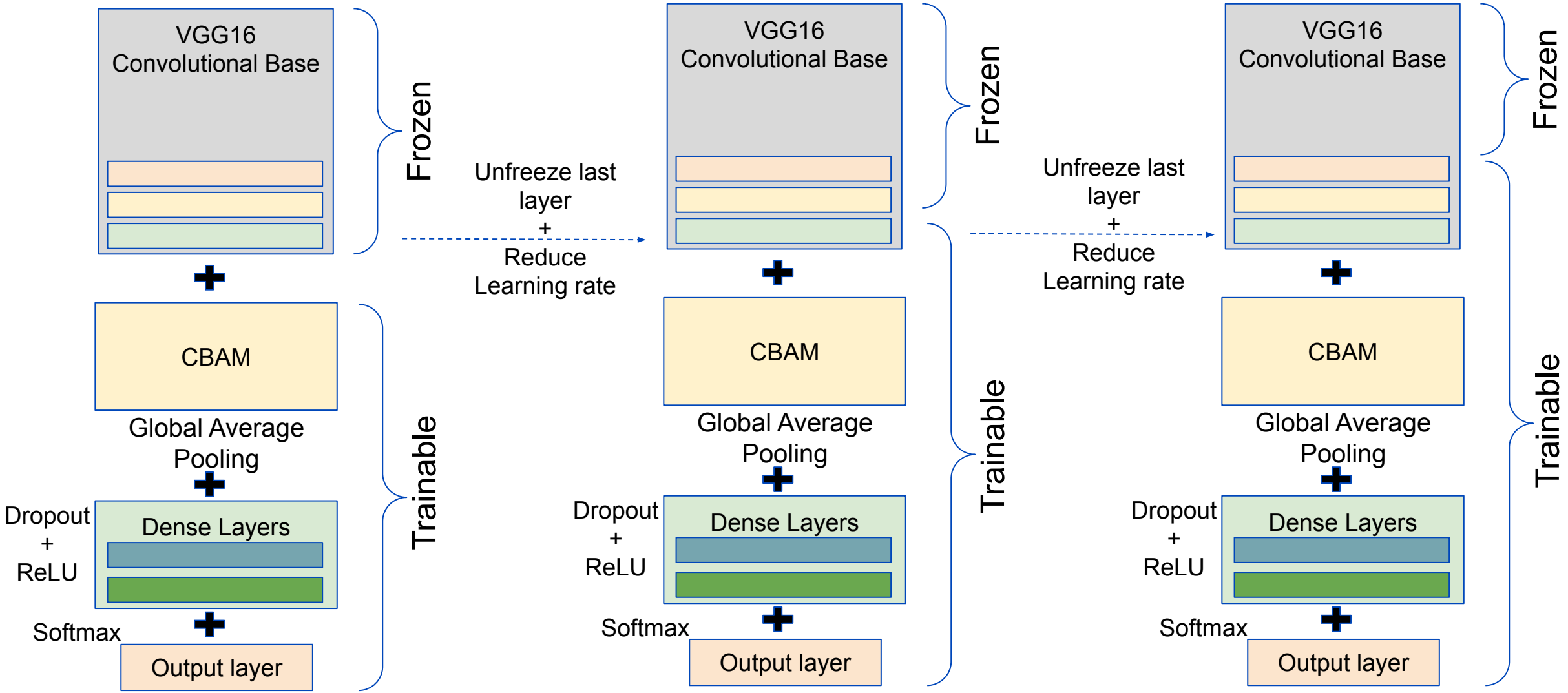
Methodology

Details of Convolutional Block Attention Module (CBAM)



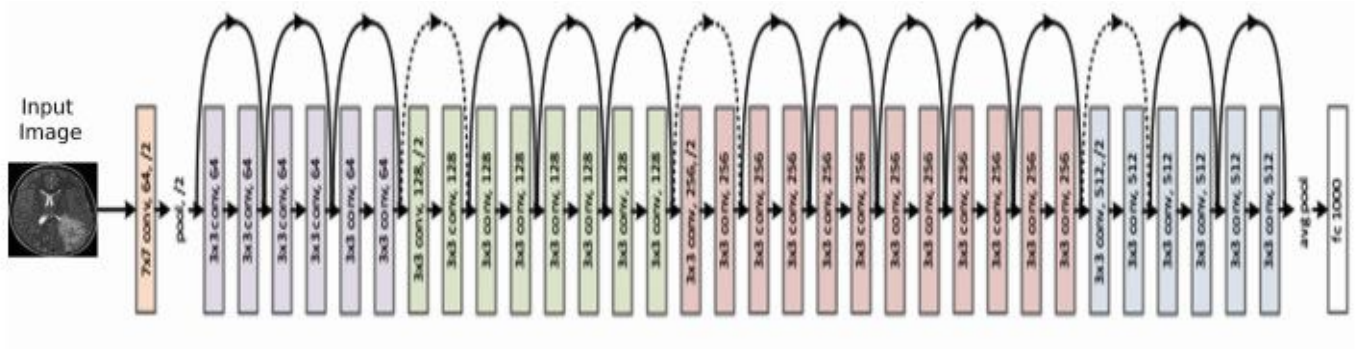
Methodology

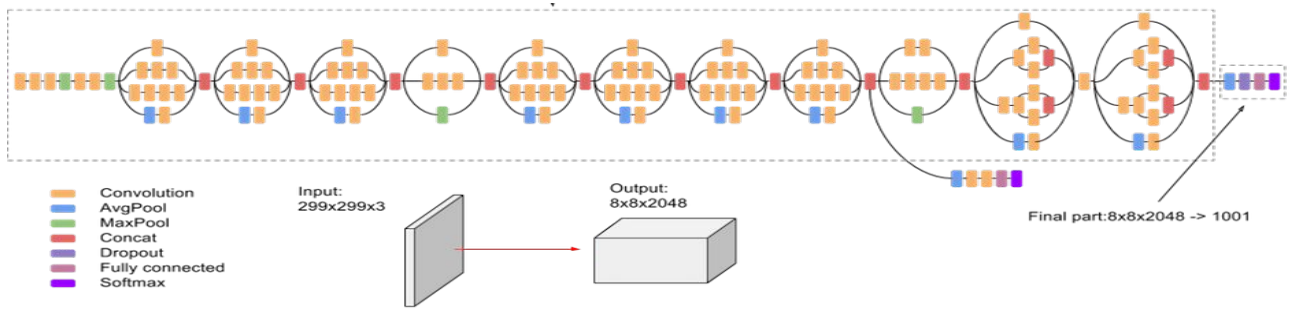
Training proposed network approach



Results & Discussion

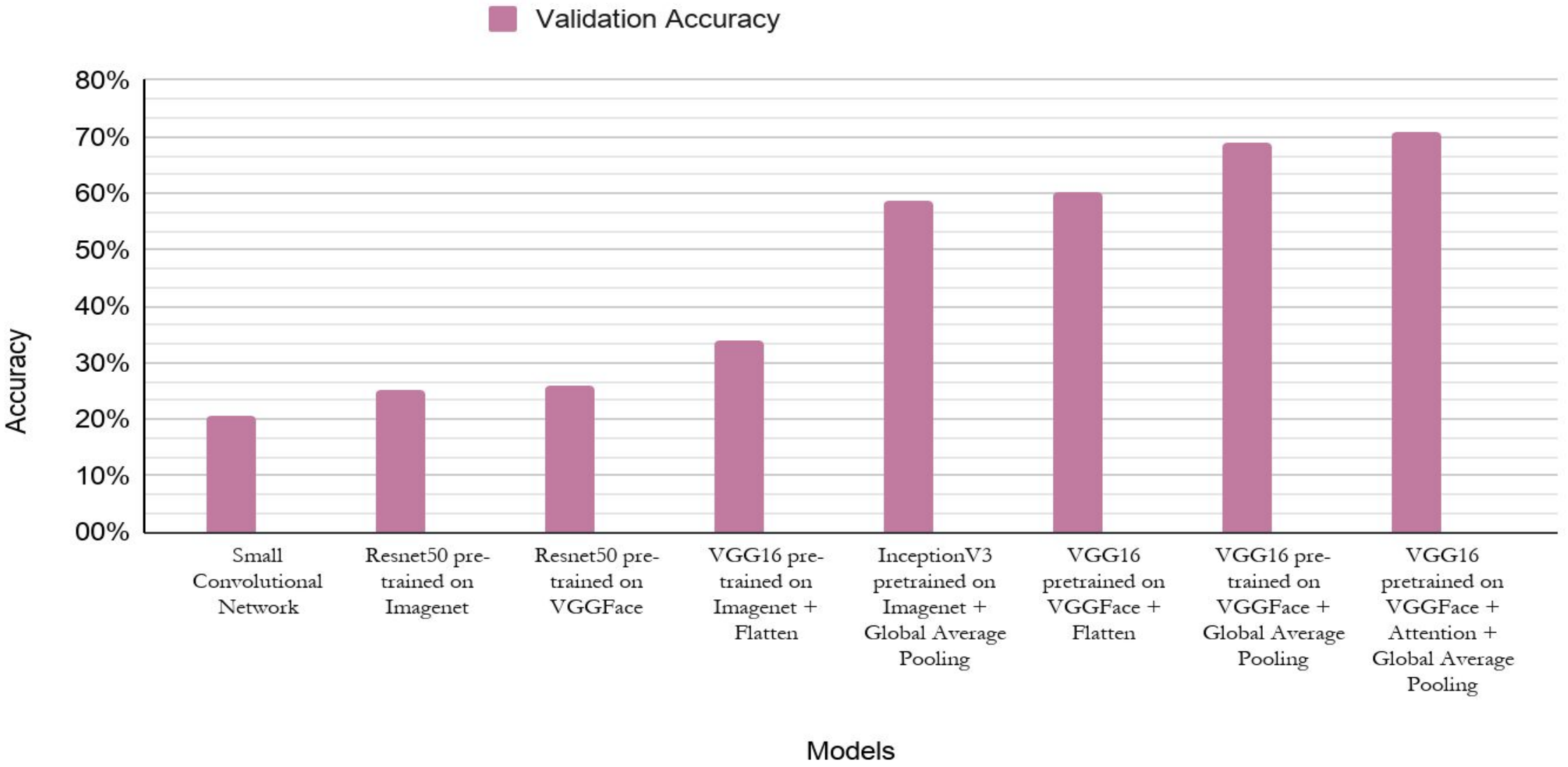
- Tested different models to improve the model with highest potential
- Models:
  - ResNet-50
  - VGG16
  - InceptionV3



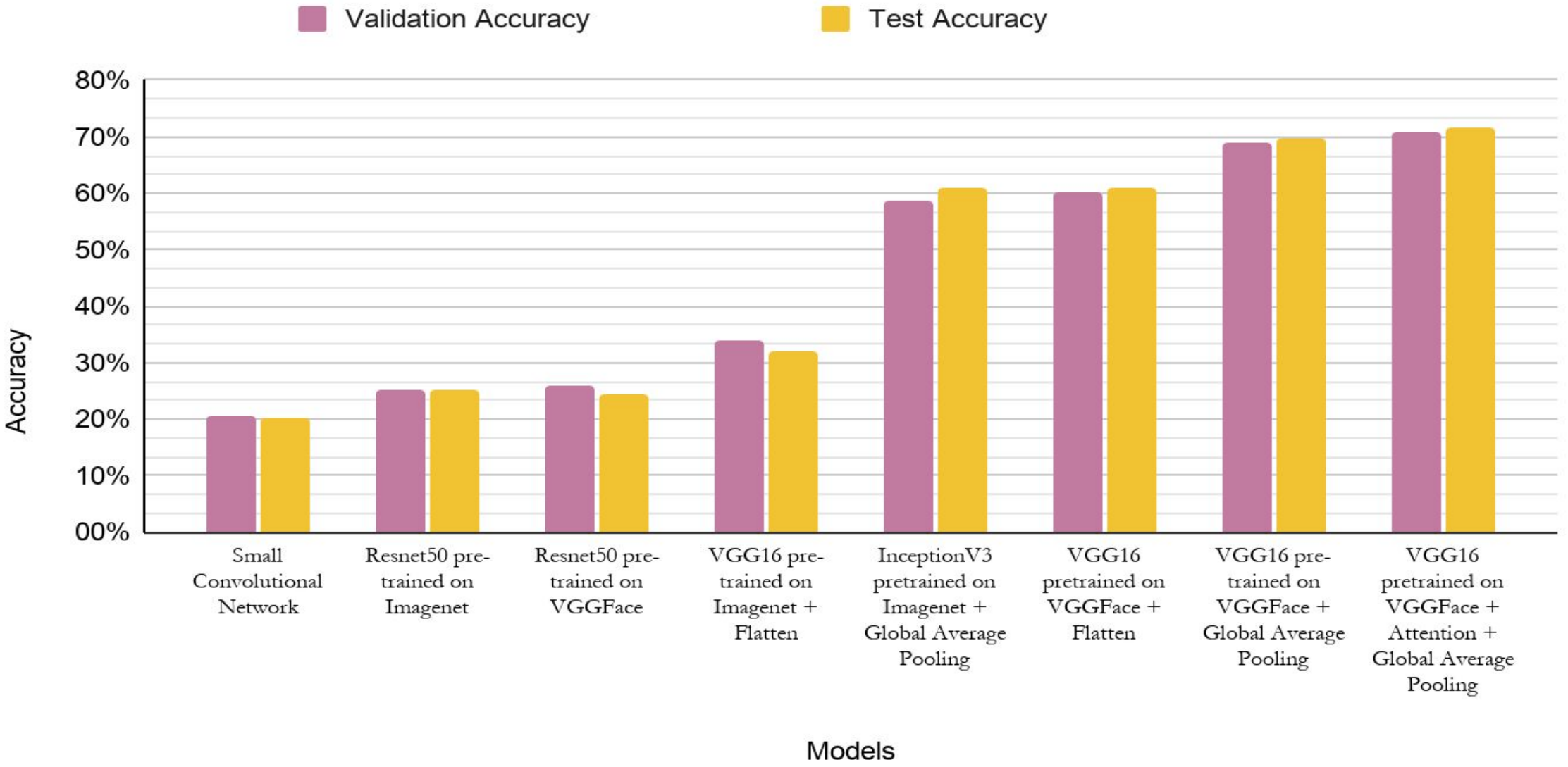


source 1: https://neurohive.io/en/popular-networks/vgg16/
source 2: https://www.aimspress.com/article/10.3934/mbe.2020328/
source 3: https://paperswithcode.com/method/inception-v3

Results & Discussion



Results & Discussion



Results & Discussion

kaggle

	Validation Accuracy (Public dataset)	Test Accuracy (Private Dataset)
Top Team in Kaggle Leaderboard	69.76 %	71.16%
Proposed Model	70.78 %	71.43 %

Source: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data

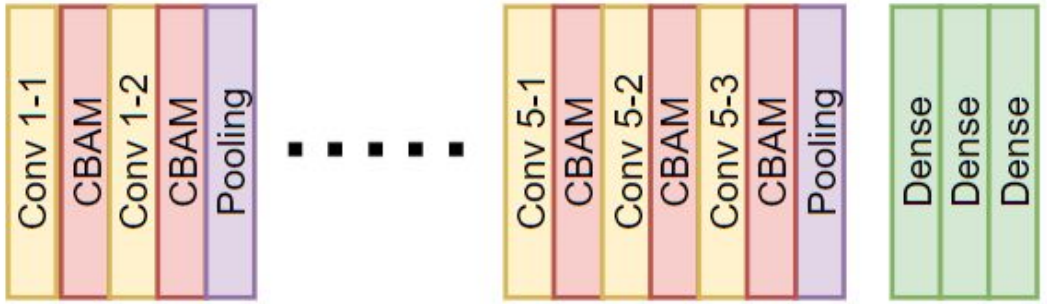
Conclusion

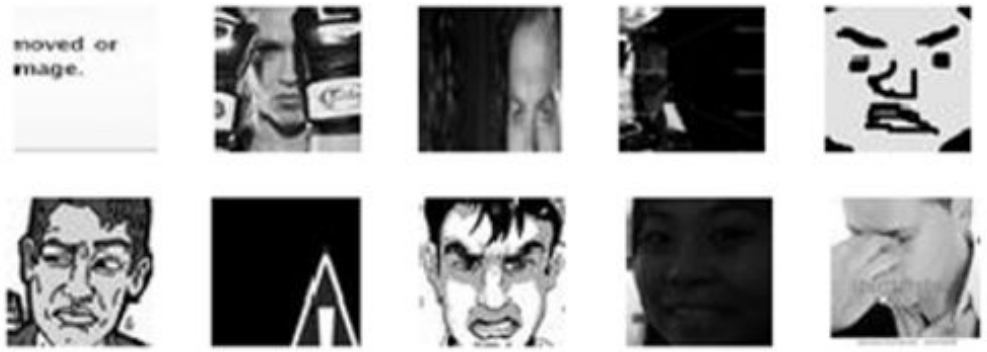
- Overcome the challenge of partial faces and occlusion
- VGG16 pretrained on VGGFace
- Convolutional Block Attention Module (CBAM)
- Dataset Problems

Pose variations, occlusion [7].

Future works

- **Attention after each convolution layer**
- **Filtering FER2013 dataset from watermarks and images include cartoon faces, not actual human faces**
- **Utilizing the Transformer network**





Source: https://github.com/amineHorseman/facial-expression-recognition-using-cnn

References

[1] M. Usman, S. Latif, and J. Qadir, “Using deep autoencoders for facial expression recognition,” in2017 13th International Conference on Emerging Technologies (ICET), pp. 1–6, 2017.

[2] F. Farooq, J. Ahmed, and L. Zheng, “Facial expression recognition using hybrid features and self-organizing maps,” in 2017 IEEE International Conference on Multimedia and Expo (ICME), pp. 409–414, 2017.

[3] S. Bellamkonda and N. Gopalan, “An enhanced facial expression recognition model using local feature fusion of gabor wavelets and local directionality patterns,”International Journal of Ambient Computing and Intelligence, vol. 11, pp. 48–70, 11 2019.

[4] S. Samsani and V. A. Gottala, “A real-time automatic human facial expression recognition system using deep neural networks,” in Information and Communication Technology for Sustainable Development(M. Tuba,S. Akashe, and A. Joshi, eds.), (Singapore), pp. 431–441, Springer Singapore, 2020.

[5] H. Siqueira, S. Magg, and S. Wermter, “Efficient facial feature learning with wide ensemble-based convolutional neural networks,”arXiv preprint arXiv:2001.06338, 2020.

[6] P. Burkert, F. Trier, M. Z. Afzal, A. Dengel, and M. Liwicki, “Dexpression:Deep convolutional neural network for expression recognition,” 2016.

[7] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, “Region attention networks for pose and occlusion robust facial expression recognition,”IEEE Transactions on Image Processing, vol. 29, pp. 4057–4069, 2020.

[8] S. Minaee and A. Abdolrashidi, “Deep-emotion: Facial expression recognition using attentional convolutional network,” 2019.

[9] S. Woo, J. Park, J.Y. Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” 2018.

جامعة خليفة
Khalifa University

Thank You

ku.ac.ae