Facemask Detection and Classification via Deep Learning

Nikhil Pathak
Joseph Chang
Benjamin Chang

Prof. Peter Gerstoft

6/3/21

UC San Diego
Electrical and Computer Engineering
JACOBS SCHOOL OF ENGINEERING

Background

Covid-19

- CDC recommends wearing face masks
- Prevents viral spread and exposure
- 83% believe face masks are effective [1]
- Only 51% practice wearing in public

Face Mask Wear

- 98% effective at preventing small particles [2]
- Many wear them incorrectly → viral spread

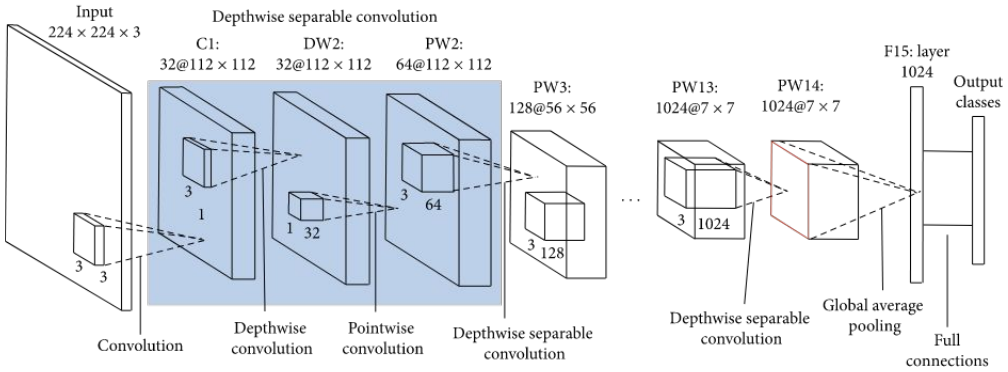
Literature Survey

- Transfer Learning [3]
  - Pretrained feature extractors
  - Flexible for solving various classification tasks
  - Reduced computational resources to training

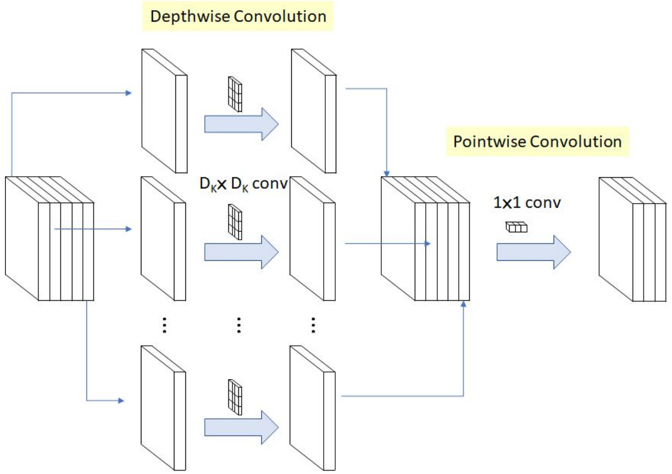
- InceptionV3 [4]
  - 48 layers deep
  - Pretrained on 1,000,000+ images
  - wide rather than deep model

- MobileNet [5]
  - Convolutional, Avg Pooling, and Dense layers
  - Depthwise separable convolutions

InceptionV3



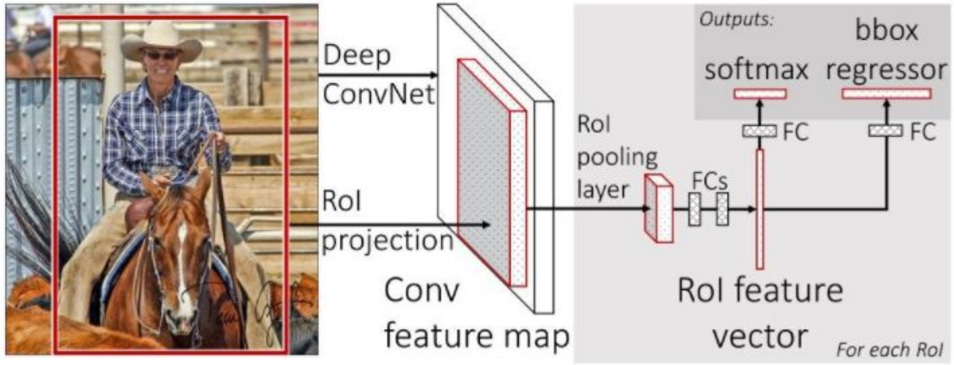
MobileNet



Literature Survey

- <u>YOLOv3</u> [6]
  - 106-layer fully convolutional architecture
  - Convolutional layers, residual layers
  - Softmax activation layer
- <u>Faster R-CNN</u> [7]
  - Convolutional layers
  - Roi pooling layer makes it fast
  - RPN (Region Proposal Generator) proposes regions for network to look at that are likely to contain objects
  - Less memory needed since doesn't cache extracted features

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
	Convolutional	32	1 × 1	
1×	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
	Convolutional	64	1 × 1	
2×	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
	Convolutional	128	1 × 1	
8×	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
	Convolutional	256	1 × 1	
8×	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
	Convolutional	512	1 × 1	
4×	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			



How can Machine Learning help?

- CNNs are great at learning information from images
  - Powerful, trainable feature extractors can help us classify whether a person is properly wearing a facemask
- Traditional approaches require more involvement from us to classify these faces
  - NN learns end-to-end

Dataset

Face Mask Detection Dataset

- Obtained from Kaggle [8]
- 853 images of different resolutions and aspect ratios
- 4072 faces with ground truth labels
- 3 Classes: Masked, Not Masked, Incorrectly Masked
- Avg Size of Face: 31.15×35 pixels → resized to 35x35 pixels
- Split into 8:1:1 training, validation, and testing images
- Used for train/val/test purposes

Dataset Imbalance Issue!

- This dataset has serious class imbalance issues!
- Training split of 3257 faces
  - 2551 Masked
  - 608 Not Masked
  - 98 Masked Incorrectly
- Traditional CrossEntropyLoss does not do well with massive class imbalance…
- Solution: Weighted CrossEntropyLoss!

$$\text{loss}(\mathbf{x}, class) = (-\mathbf{w}[class]) \log \left(\frac{e^{\mathbf{x}[class]}}{\sum_j e^{\mathbf{x}[j]}} \right)$$

$$\text{loss for batch size } N = \frac{\sum_{i=1}^{N} \text{loss}(\mathbf{pred}[i], class[i])}{\sum_{i=1}^{N} \mathbf{w}[class[i]]}$$

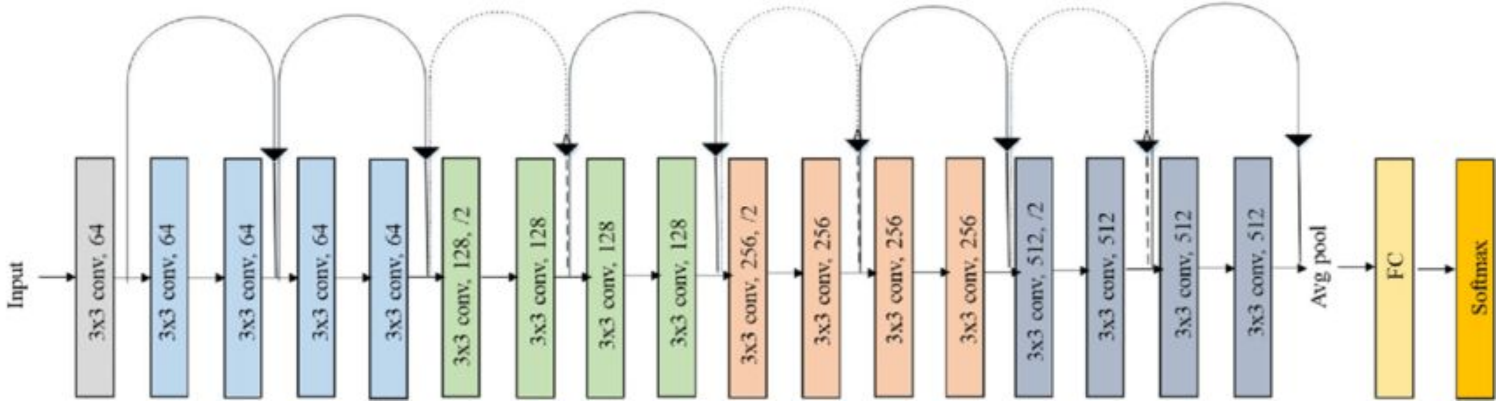
Masked = 98/2551
w = Unmasked = 98/608
Masked Incorrectly = 98/98

Models

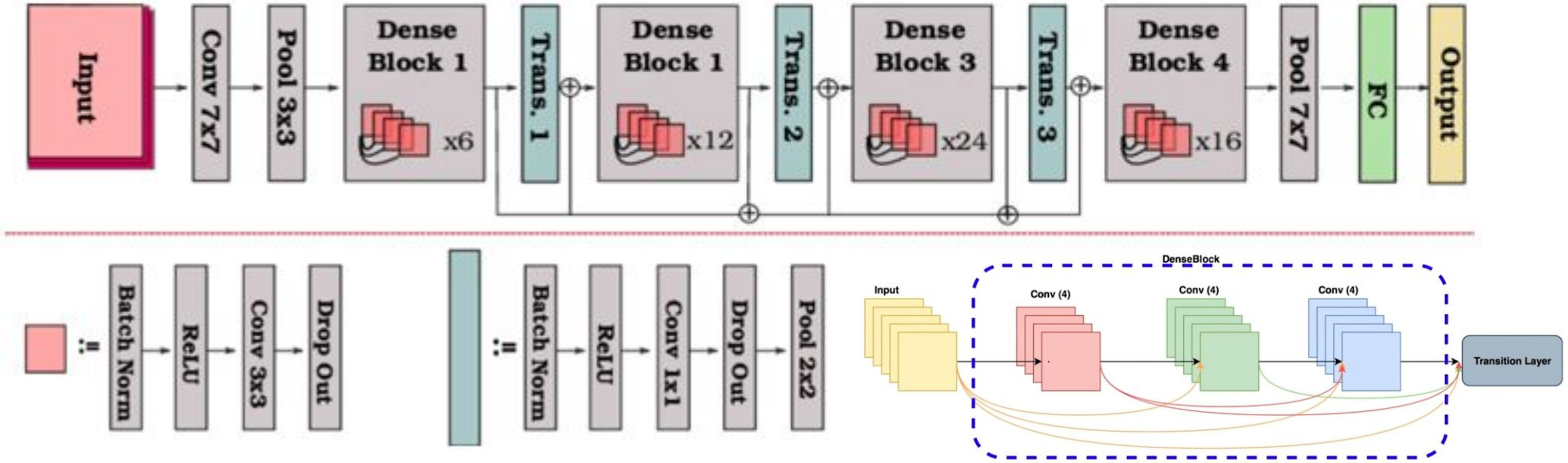
- Transfer Learning
- Weighted CrossEntropyLoss to train models
- Metric to Evaluate Models: Average-Class-Accuracy (ACA)
  - ACA = (Class1 Acc. + Class2 Acc. + Class3 Acc.) / 3
- Our project explored alternative Transfer Learning models including:
  - ResNet-34 [8]
  - VGG-16 [9]
  - VGG-19
  - DenseNet201
  - DenseNet161
  - DenseNet121
- Also experimented with a Custom CNN architecture
  - BatchNormalization
  - ReLU vs. TanH activation functions

Model Details

ResNet [9]
- Varying layers (i.e. 18, 50, 152)
- Skip connections to deal with vanish/exploding gradient

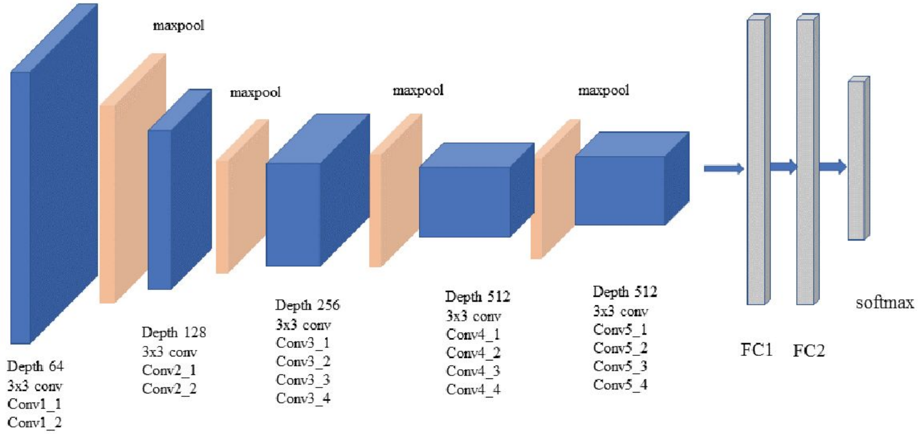


DenseNet [10]
- Every layer connected to previous layers in Dense Blocks
- Every layer adds limited parameters (12 kernels learnt per layer)

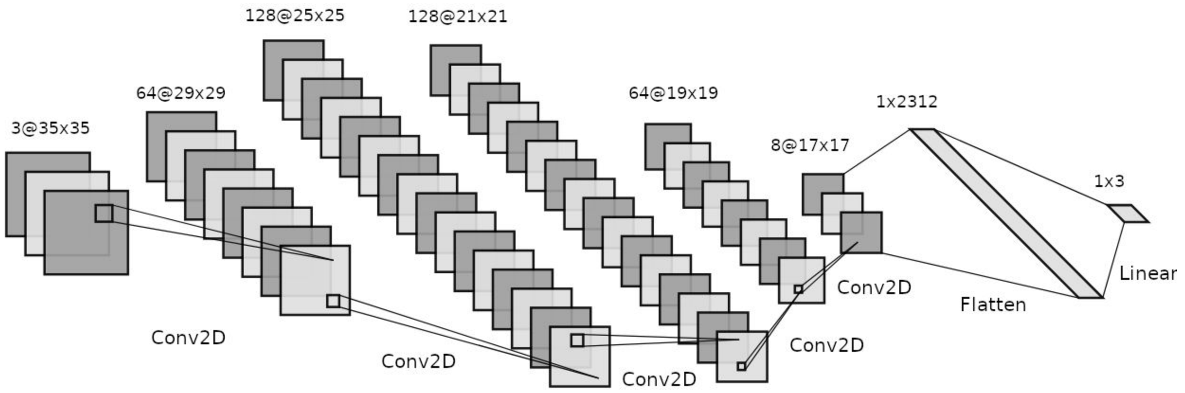


Model Details

VGG [11]
- Deep architecture with "very small (3x3)" convolution filters
- Trained on ImageNet database
- Uses Convolutional, Pooling, and Dense layers



Custom
- Convolution filters
- Activations: ReLU, TanH
- Layers: BatchNorm



Results

<u>Custom CNN with ReLU + BatchNorm</u>

- Best Class Accuracy: “Mask”
- Worst Class Accuracy: “Inc. Mask”
- Average Class Accuracy (ACA): 85.29%

<u>VGG16</u>

- Best Class Accuracy: “Mask”
- Worst Class Accuracy: “Inc. Mask”
- Average Class Accuracy (ACA): 83.60%

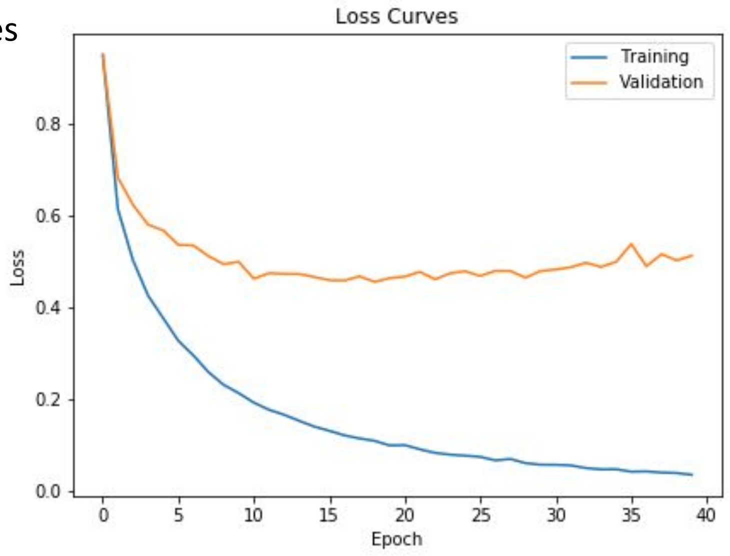
<u>VGG19</u>

- Best Class Accuracy: “No Mask”
- Worst Class Accuracy: “Inc. Mask”
- Average Class Accuracy (ACA): 83.23%

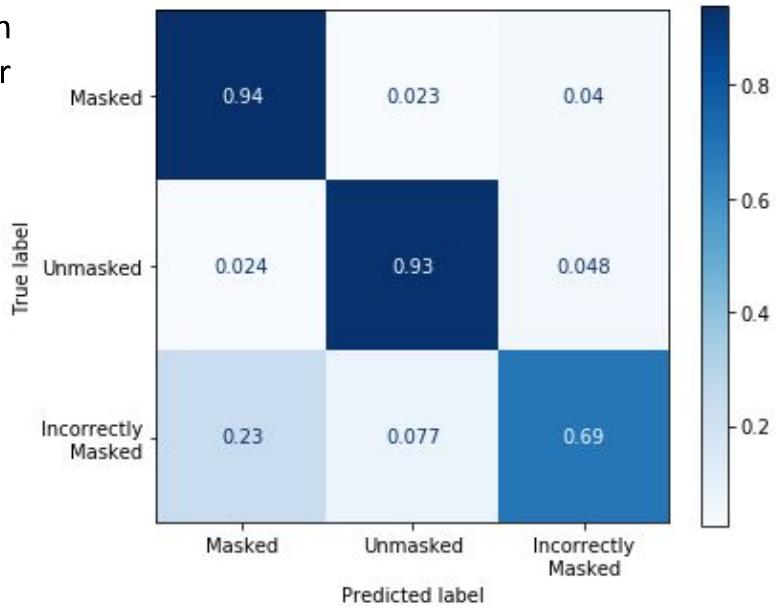
	Accuracies			
Model	Mask	No Mask	Inc. Mask	ACA
ResNet18	96.03	80.95	23.08	**66.69**
ResNet18 w/ 2 FC Layers	96.03	83.33	15.39	**64.92**
ResNet34	89.80	90.48	30.77	**70.35**
Custom CNN with ReLU	90.94	88.10	46.15	**75.06**
Custom CNN with ReLU + BatchNorm	93.77	92.86	69.23	***<u>85.29</u>***
Custom CNN with TanH	86.69	90.48	53.85	**77.01**
Custom CNN with TanH + BatchNorm	86.97	95.24	46.15	**76.12**
VGG16	93.48	88.10	69.23	**83.60**
VGG16 with BatchNorm	90.94	71.43	24.18	**62.18**
VGG19	92.92	95.24	61.54	**83.23**
VGG19 with BatchNorm	90.94	85.71	30.77	**69.14**
DenseNet201	98.87	95.24	23.10	**72.40**
DenseNet161	98.58	95.24	46.15	**79.99**
DenseNet121	95.75	90.48	30.77	**72.33**

Best Model (Custom CNN with ReLU + BatchNorm) Results

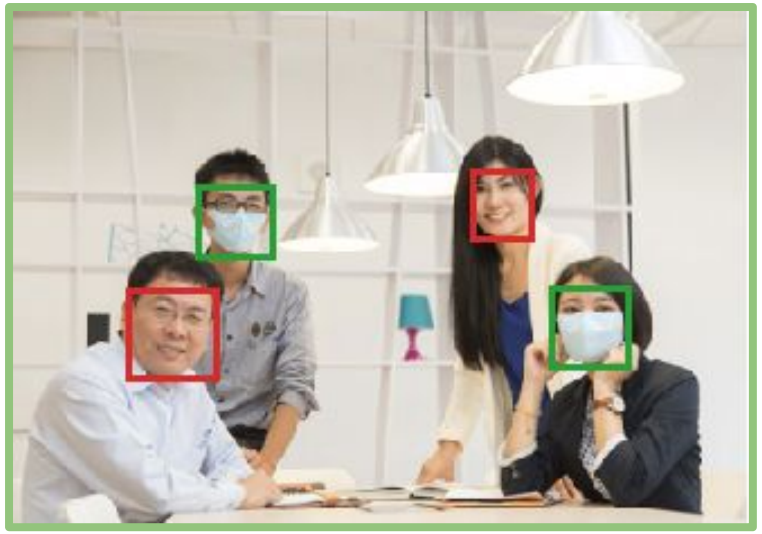
Loss Curves for Best Model



Confusion Matrix for Test Set

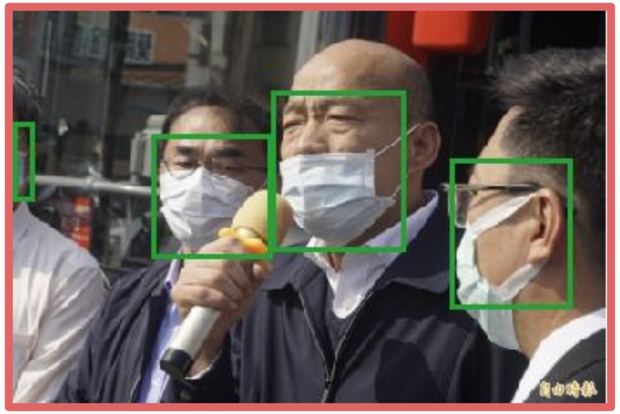


Good Predictions:



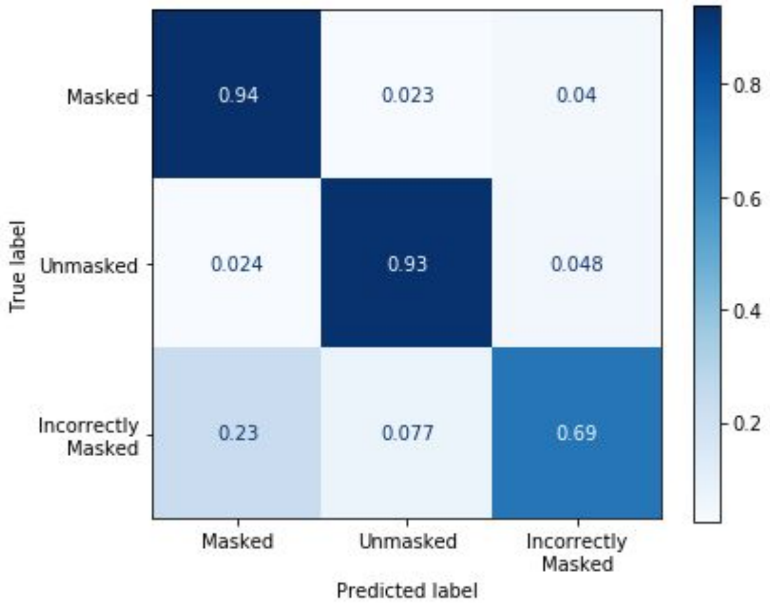
Key for Predictions:
Masked
Unmasked
Masked Incorrectly

Bad Predictions:



Observations

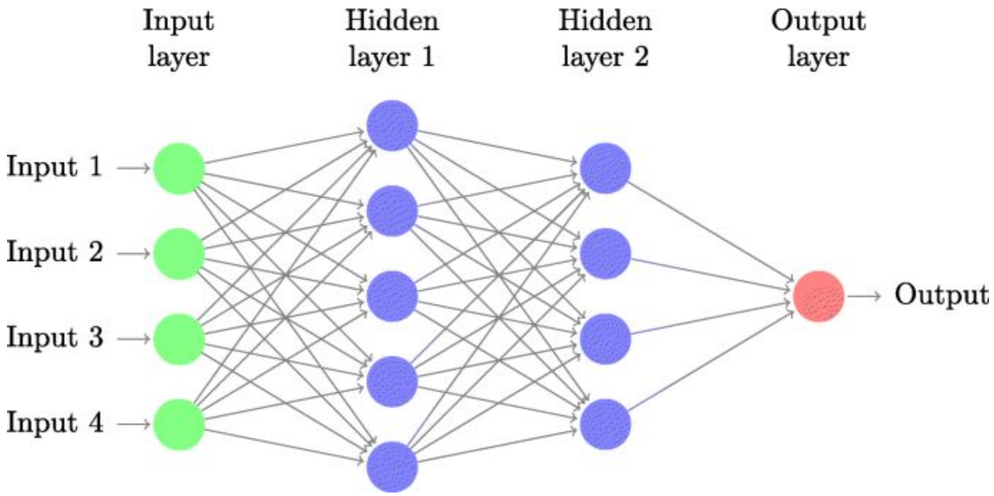
- Custom CNN with ReLU + BatchNorm model has highest accuracy
- Potential to train more accurate face mask classification models



Custom CNN Confusion Matrix

Future Work

- Explore using Faster R-CNN model directly
- Increase accuracy on “incorrectly masked faces” class by collecting more data to train on
- Test with more model layers



References

[1] Benjamin Fearnow, “Americans Support Wearing Masks, But Only Half Wear Them,” in Newsweek, Jan. 23, 2021, https://www.newsweek.com/83-percent-americans-support-wearing-masks-only-half-wear-them-poll-1563944.

[2] Brenda Goodman, “How Much Does Wearing a Mask Protect You?” in WebMD, Nov. 19, 2020, https://www.webmd.com/lung/news/20201119/how-much-does-wearing-a-mask-protect-you.

[3] Wikipedia, “Transfer learning - Wikipedia, the free encyclopedia,“ *http://en.wikipedia.org/w/index.php?title=Transfer\%20learning&oldid=1022394104*, 2021.

References

[4] G. Jgnesh Chowdary, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal, “Face Mask Detection using Transfer Learning of InceptionV3,” 2009.

[5] Isunuri Venkateswarlu, Jagadeesh Kakarla, and Shree Prakash, “Face mask detection using MobileNet and Global Pooling Block,” 2020.

[6] Joseph Redmon and Ali Farhadi, “Yolov3: An incremental improvement,” 2018.

[7] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” 2016.

References

[8] Larxel, “Face Mask Detection Dataset” in Kaggle, 2020, https://www.kaggle.com/andrewmvd/face-mask-detection.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” 2015.

[10] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger, “Densely Connected Convolutional Networks,” 2018.

[11] Nagesh Singh Chauhan, “Mask and social distancing detection using vgg19,” *https://www.kaggle.com/nagehsingh/mask-and-social-distancing-detection-using-vgg19*, Feb 2021.

Thank you!