Exercise 3.2
Image classification

Group 1

Mehdin Masinovic (Student ID: 11713301)
Fardokhtsadat Mohammadi (Student ID: 01655989)
Waldemar Schewzow (Student ID: 12102485)

Task description

- Color histograms
- SIFT + Visual Bag of Words
- CNN

Datasets

Fashion MNIST:

- Provided by Zalando as grey images (60000 train, 10000 test)
- To replace old handwritten-digits dataset, more difficult to extract features
- 10 classes: shoes, pants, skirts, ...

Labelled faces in the wild:

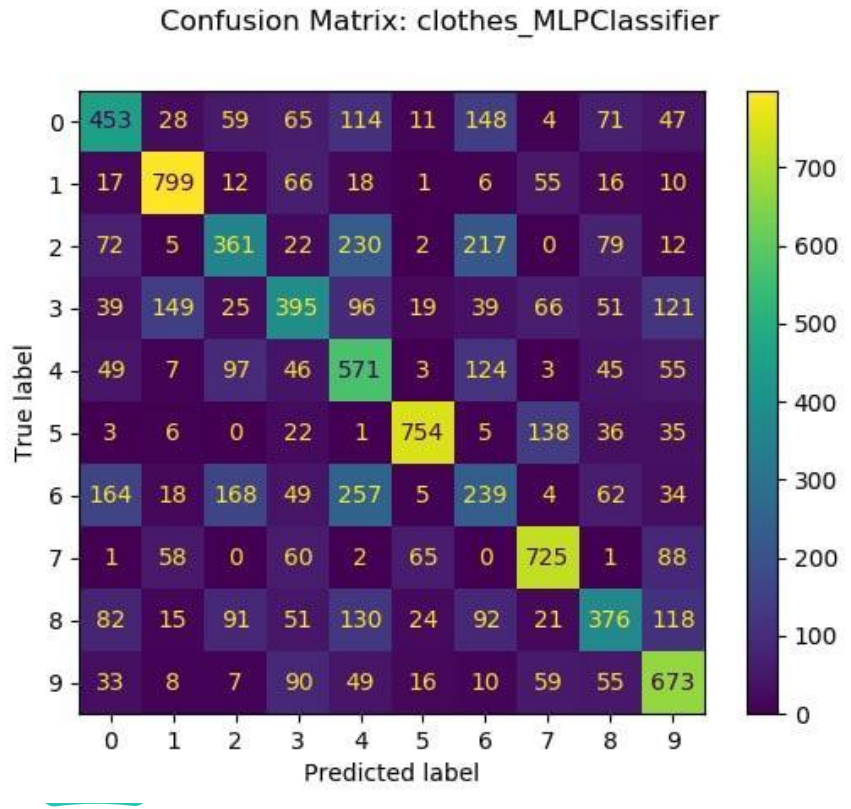
- Provided by scikit-learn (3024 rows)
- Only people used that have at least 20 images
- Used to recognize different people, every person is one class (42 classes)

Color Histograms

Idea:

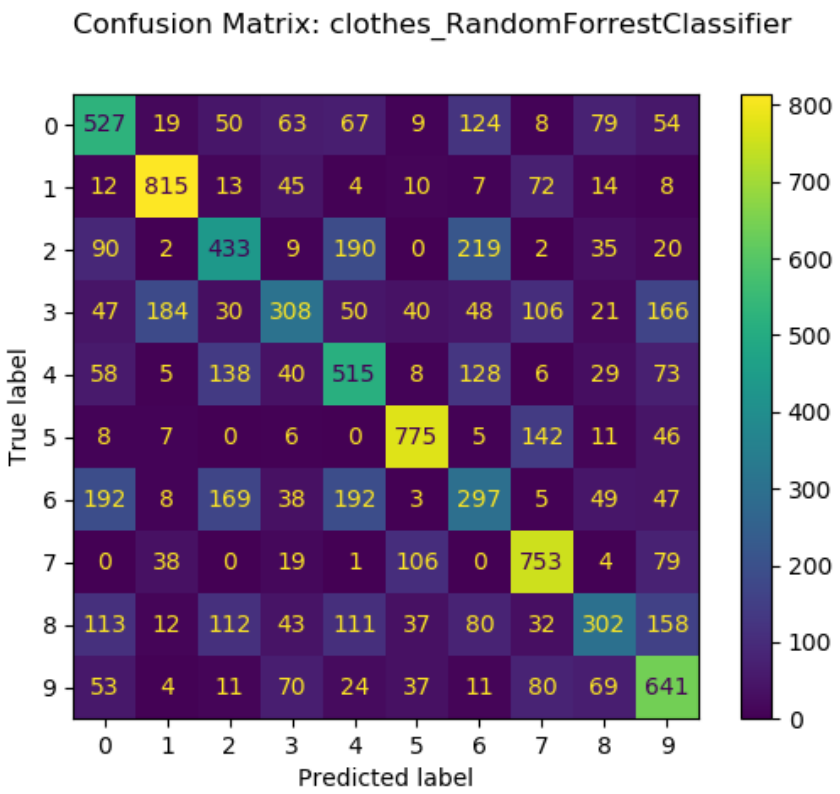
- For every image, use brightness in pixels to compute histogram (0-255)
- Grey images have one channel (0=black, 255=white), whereas colorful images have three (red, green, and blue)
- Split data into training and test sets
- Train models, apply on the test set, and measure and compare performance of models

Color Histogram – Fashion MNIST – MLP – Activation Relu – Learning Constant – 3 Layers with 50, 100 and 50 Neurons



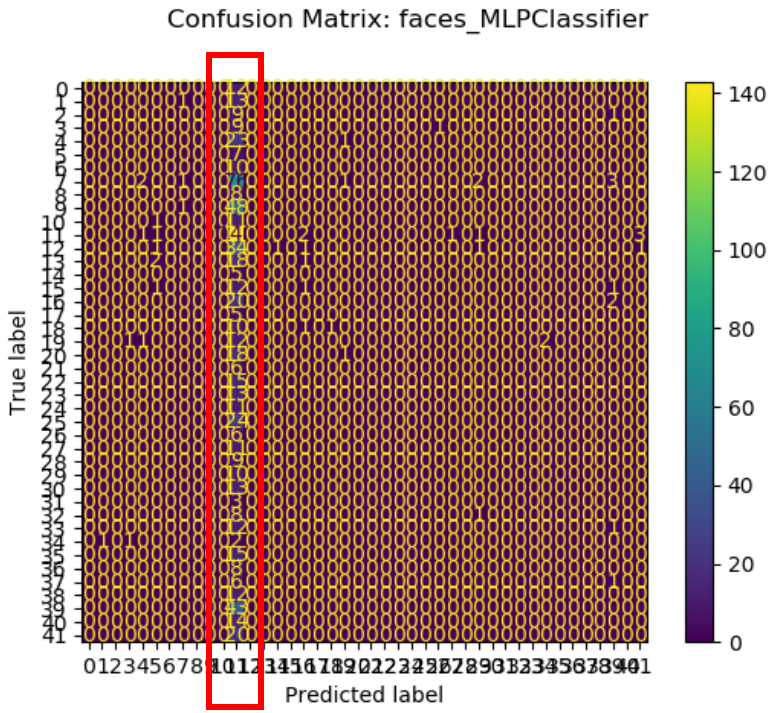
Class Name	Precision	Recall	F1-score
T-shirt/top	0.50	0.453	0.47
Trouser	0.73	0.799	0.76
Pullover	0.44	0.361	0.40
Dress	0.46	0.395	0.42
Coat	0.39	0.571	0.46
Sandal	0.84	0.754	0.80
Shirt	0.27	0.239	0.25
Sneaker	0.67	0.725	0.70
Bag	0.47	0.376	0.42
Ankle boot	0.56	0.673	0.61
Accuracy	0.53	0.5346	0.53

Color Histogram – Fashion MNIST – Random Forest – Max Depth 20 – Nr of Estimators 100



Class Name	Precision	Recall	F1-score
T-shirt/top	0.48	0.53	0.50
Trouser	0.74	0.81	0.78
Pullover	0.45	0.43	0.44
Dress	0.48	0.31	0.37
Coat	0.45	0.51	0.48
Sandal	0.76	0.77	0.76
Shirt	0.32	0.30	0.31
Sneaker	0.62	0.75	0.69
Bag	0.50	0.30	0.37
Ankle boot	0.50	0.64	0.56
Accuracy	0.54	0.54	0.5366

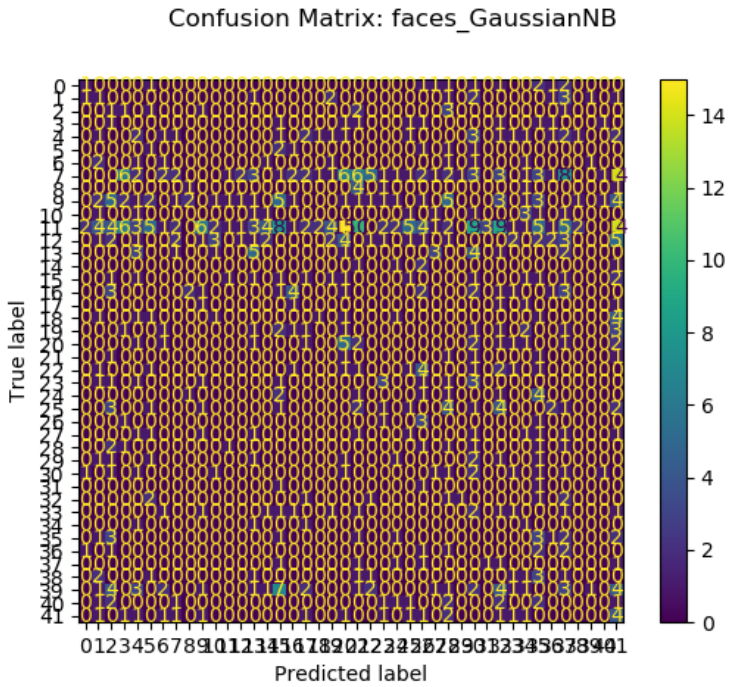
Color Histogram – Labelled Faced in The Wild – MLP – Activation Relu – Learning Constant – 1 Layer with 100 Neurons



Class Name	precision	recall	f1-score	support
Colin Powell	0.33	0.012	0.02	85.0
George W. Bush	0.19	0.94	0.31	152.0
Jack Straw	1.0	0.083	0.15	12.0
accuracy	0.18	0.18	0.18	0.18

George W. Bush: 530
Colin Powell: 236
Jack Straw: 28

Color Histogram – Labelled Faced in The Wild – Naive Bayes with Default Parameters



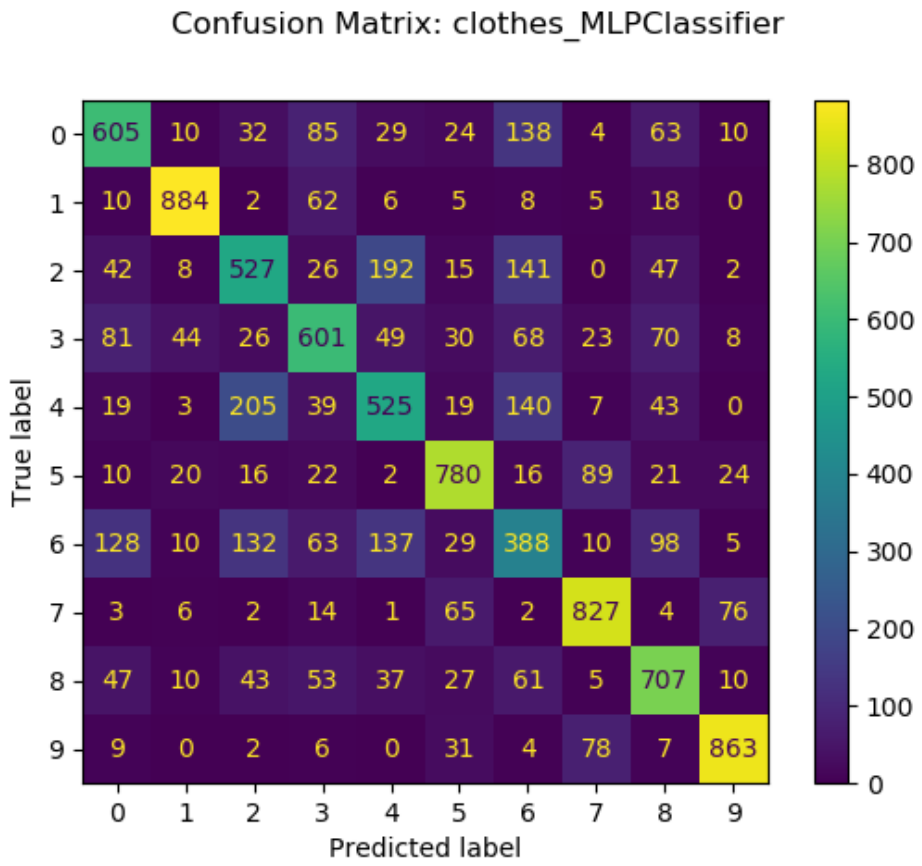
Class Name	Precision	Recall	F1-score
0	0.17	0.08	0.11
1	0.04	0.07	0.05
2	0.03	0.1	0.04
4	0.08	0.08	0.08
5	0.05	0.14	0.08
6	0.10	0.1	0.09
7	0.18	0.02	0.04
8	0.17	0.12	0.14
11	0.2	0.006	0.01
12	0.2	0.02	0.04
13	0.22	0.23	0.22
15	0.03	0.07	0.04
16	0.31	0.18	0.23
18	0.2	0.08	0.12
20	0.12	0.263	0.16
23	0.23	0.23	0.23
25	0.06	0.04	0.05
26	0.10	0.5	0.17
28	0.03	0.11	0.05
30	0.04	0.15	0.06
33	0.08	0.077	0.08
34	0.05	0.25	0.087
35	0.07	0.2	0.10
39	0.33	0.02	0.04
41	0.06	0.2	0.09
Accuracy	0.06	0.06	0.06

SIFT and BOVW

Idea:

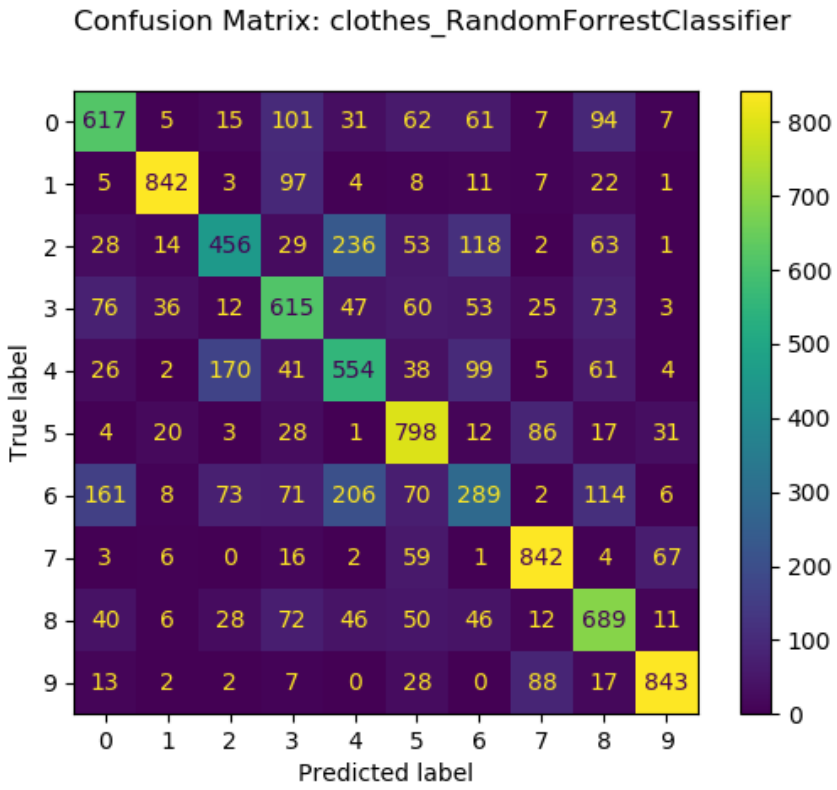
- Transform pixel information into descriptors
- Descriptors offer a numerical representation of the image, which is invariant to the rotation, size, or angle of the image
- Descriptors are clustered into similar features using K-means

SIFT – Fashion MNIST – MLP – Activation Relu – Learning Constant – 1 Layer with 100 Neurons



Class Name	Precision	Recall	F1-score
T-shirt/top	0.63	0.60	0.62
Trouser	0.89	0.88	0.89
Pullover	0.53	0.53	0.53
Dress	0.62	0.60	0.61
Coat	0.54	0.52	0.53
Sandal	0.76	0.78	0.77
Shirt	0.40	0.39	0.39
Sneaker	0.79	0.83	0.81
Bag	0.65	0.71	0.68
Ankle boot	0.86	0.86	0.86
Accuracy	0.67	0.67	0.67

SIFT – Fashion MNIST – Random Forest – Max Depth 20 –Nr of Estimators 100

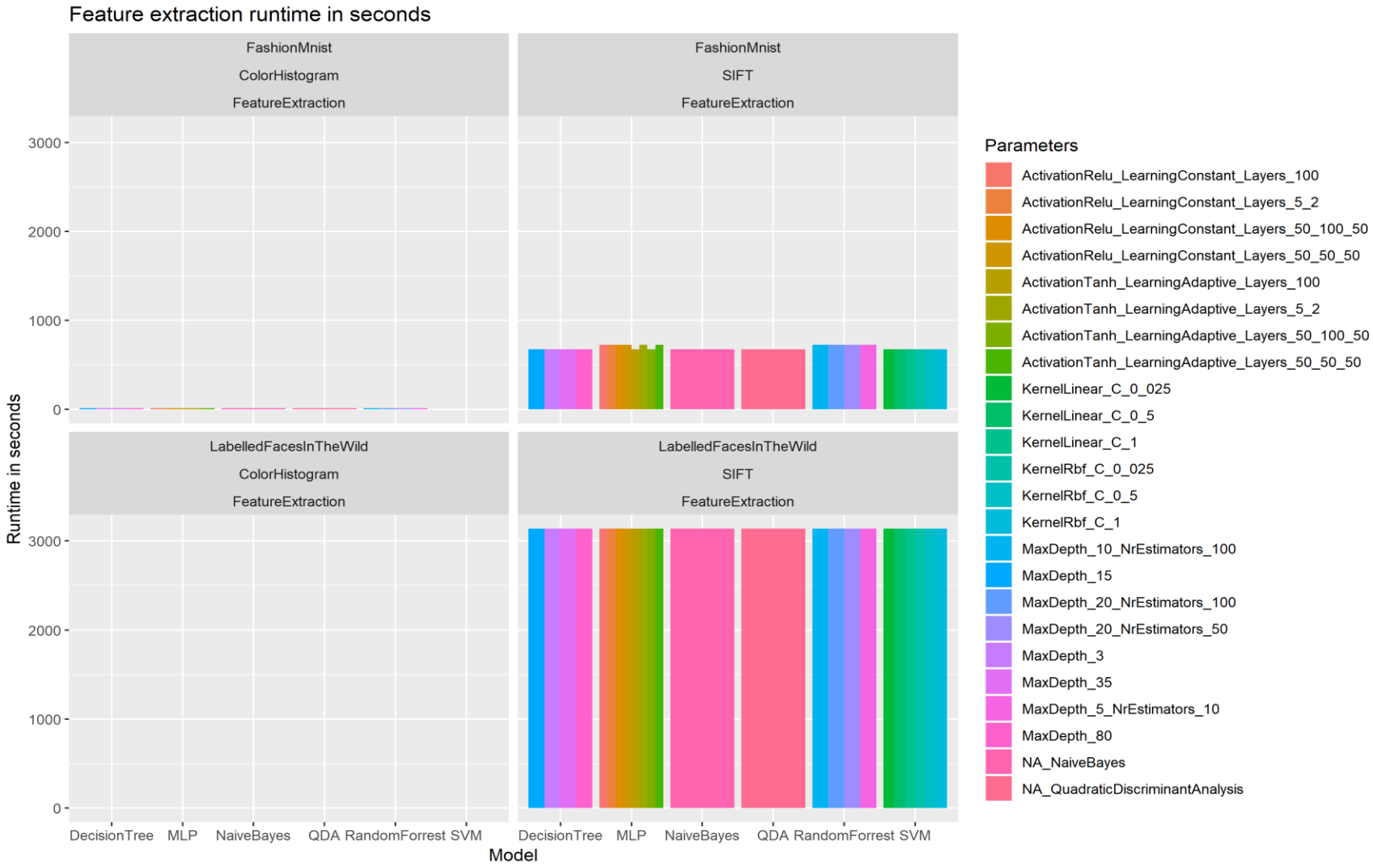


Class Name	Precision	Recall	F1-score
T-shirt/top	0.63	0.62	0.62
Trouser	0.90	0.84	0.87
Pullover	0.60	0.46	0.52
Dress	0.57	0.61	0.60
Coat	0.50	0.55	0.52
Sandal	0.65	0.80	0.72
Shirt	0.42	0.29	0.34
Sneaker	0.78	0.84	0.81
Bag	0.60	0.69	0.64
Ankle boot	0.86	0.84	0.85
Accuracy	0.65	0.65	0.65

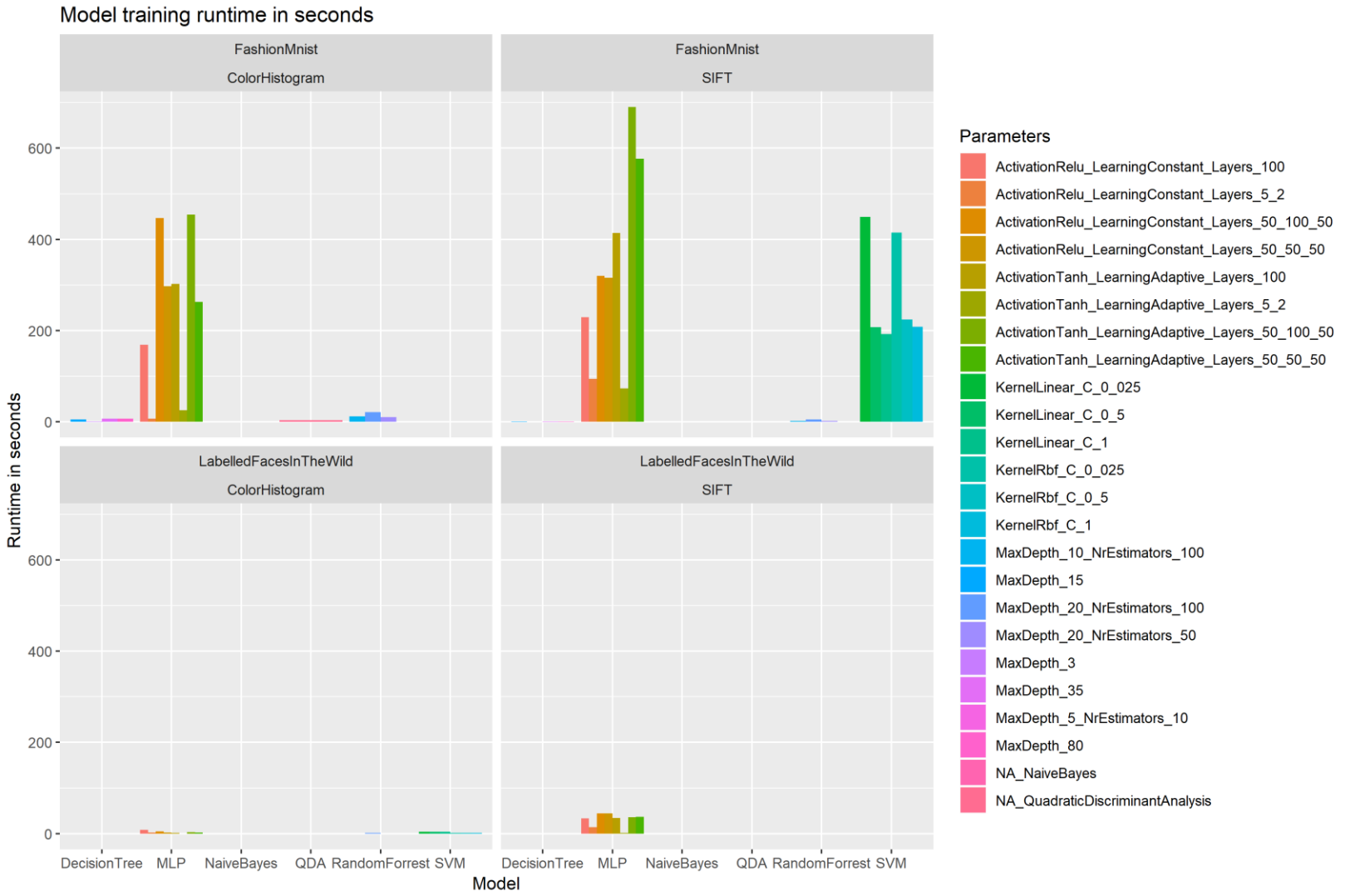
SIFT – Labelled Faced in The Wild

- Similar results to the color histogram approach:
  - Most models predict majority class only
  - Naïve Bayes offers best classification results

Time and Performance for Feature Extraction



Time and Performance for Model Training



CNN Architectures

- MiniVGGNet
- MiniGoogLeNet

Setup

Fashion MNIST				*LFW Faces*			
Without augmentation		*With augmentation*		*Without augmentation*		*With augmentation*	
MVGGN	*MGLN*	*MVGGN*	*MGLN*	*MVGGN*	*MGLN*	*MVGGN*	*MGLN*
LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4
LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3
LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2
LR 1e-1	LR 1e-1	LR 1e-1	LR 1e-1	LR 1e-1	LR 1e-1	LR 1e-1	LR 1e-1

Setup

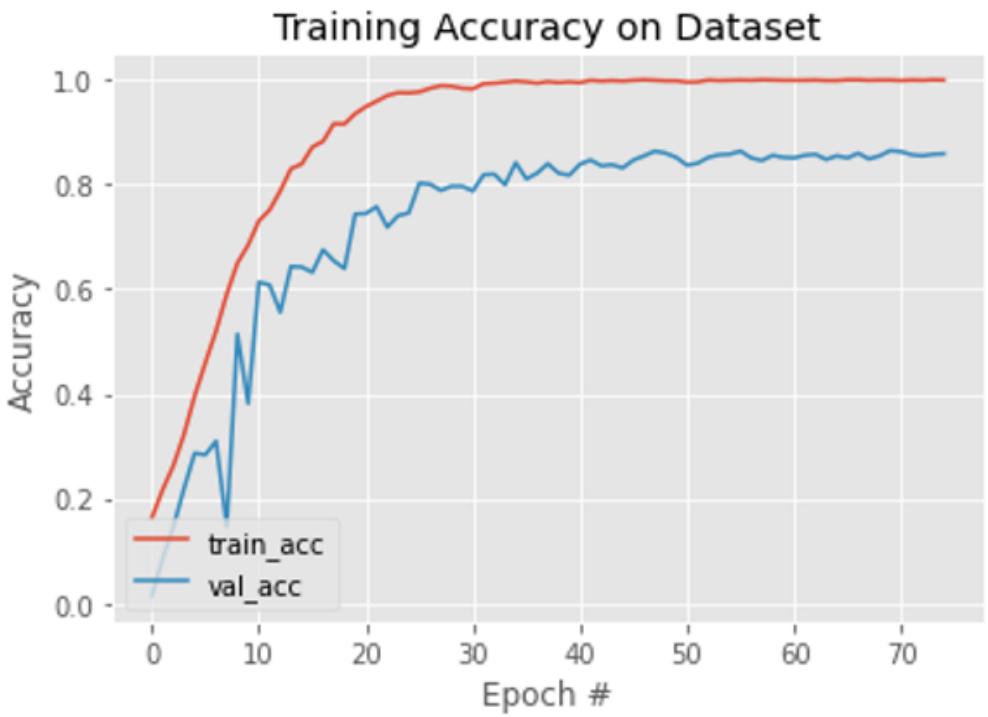
Fashion MNIST				*LFW Faces*			
Without augmentation		*With augmentation*		*Without augmentation*		*With augmentation*	
MVGGN	*MGLN*	*MVGGN*	*MGLN*	*MVGGN*	*MGLN*	*MVGGN*	*MGLN*
LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4	LR 1e-4
LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3	LR 1e-3
LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2	LR 1e-2
LR 1e-1	LR 1e-1	LR 1e-1	LR 1e-1	LR 1e-1	**LR 1e-1**	LR 1e-1	**LR 1e-1**

accuracy			0.88
macro avg	0.87	0.83	0.83
weighted avg	0.90	0.88	0.88
	(Precision)	(Recall)	(F1)

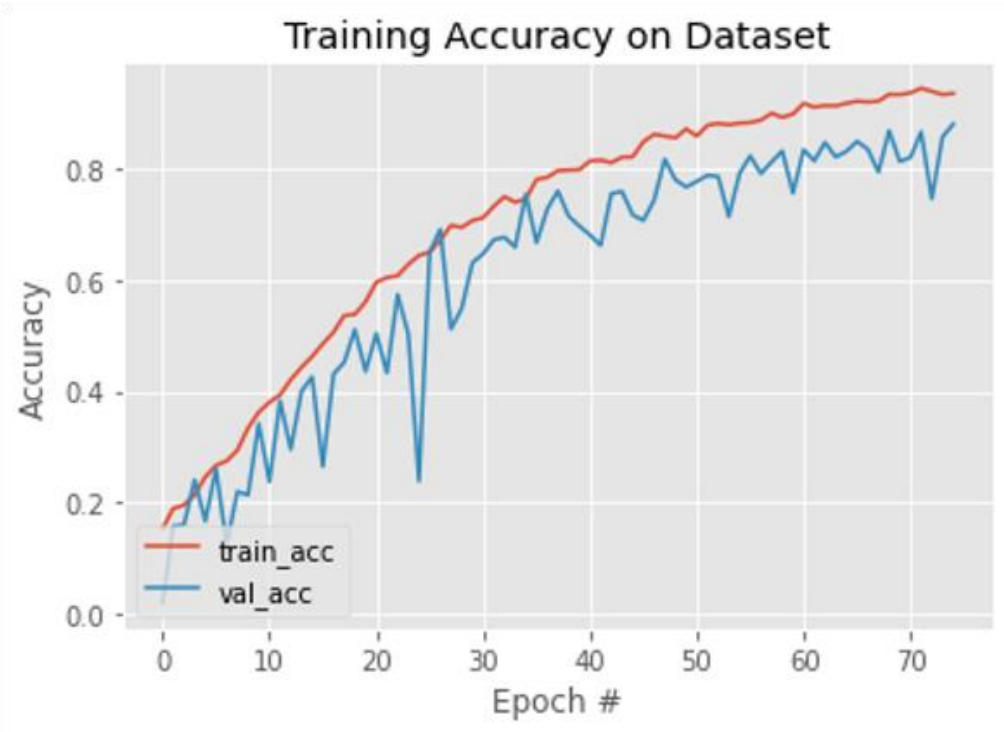
accuracy			0.86
macro avg	0.84	0.82	0.82
weighted avg	0.87	0.86	0.86
	(Precision)	(Recall)	(F1)

accuracy			0.88
macro avg	0.87	0.83	0.83
weighted avg	0.90	0.88	0.88
	(Precision)	(Recall)	(F1)

Augmentation and learning curves

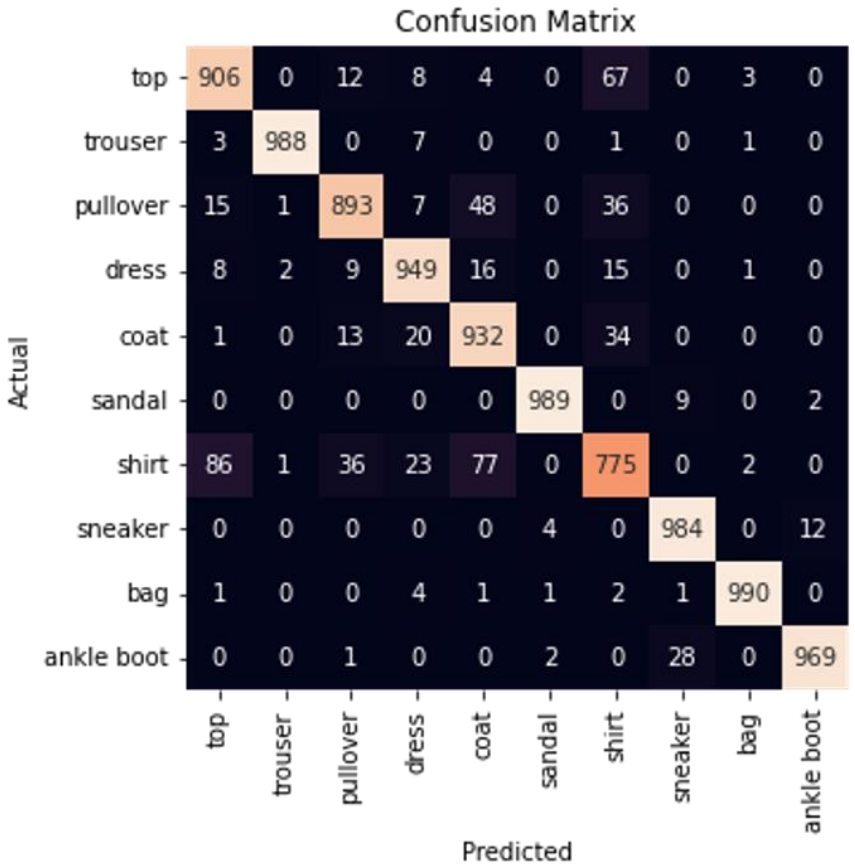


Without augmentation

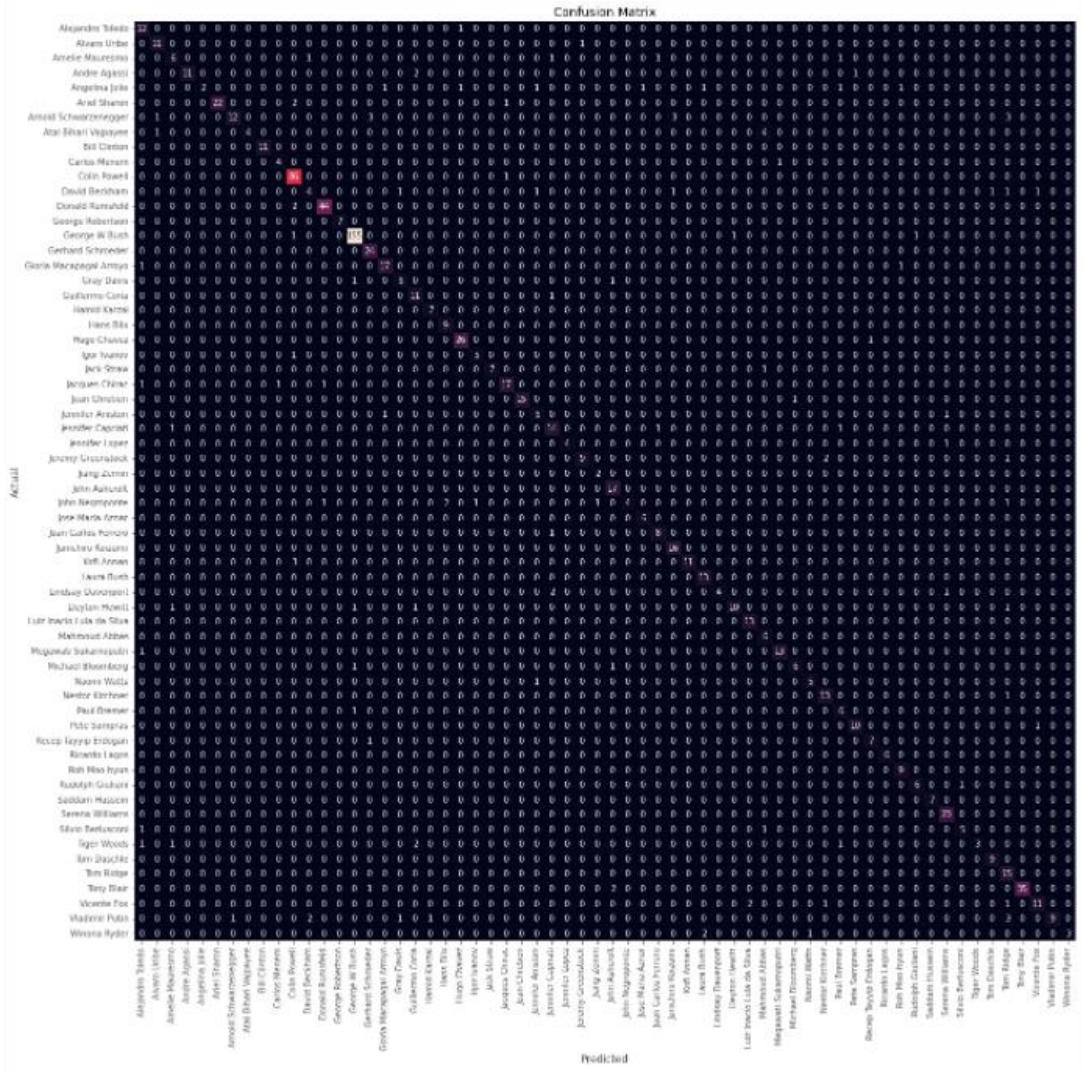


With augmentation

CNN – Fashion MNIST - MiniVGGNet



CNN – Labelled Faced in The Wild - MiniGoogLeNet



Conclusion

Color histograms:

- Very fast runtime
- Simple implementation
- Considering simplicity, very good results

SIFT + BoVW:

- More computation, but can be parallelized
- Features can be controlled using Bag of Visual Words in K-means clustering
- Implementation is more difficult, requires deeper understanding & knowledge
- Superior results compared to color histograms

Conclusion

MiniVGGNet

More lightweight model

Outperformed with Fashion MNIST dataset

MiniGoogLeNet

More heavyweight model

Outperformed with LFW people dataset

Required minimum image size

Challenges and Limitations

Traditional:

- SIFT approach often was not able to retrieve valuable information from image -> exception handling needed to loosen thresholds
- MLP would not converge for smaller number of iterations (up to ~200), was set to 1000.
- Computation of SVM took the longest among all methods

Deep learning:

- Resource limitations
- To get good results, long training was necessary
- Difficulties setting seed for Keras