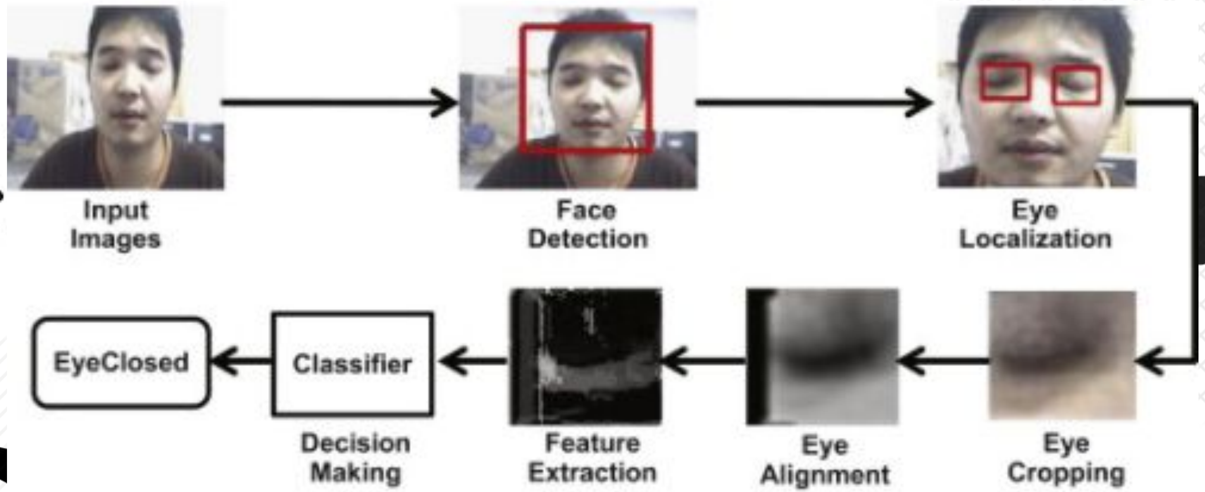
The Eye and Pupil Tracking and Segmentation with Gaze Estimation using RGB images

Aldiyar Bolatov, 201536038
Asset Rayev, 201575338
Agakhan Baiturov, 201513666
Timur Tassov, 201512793

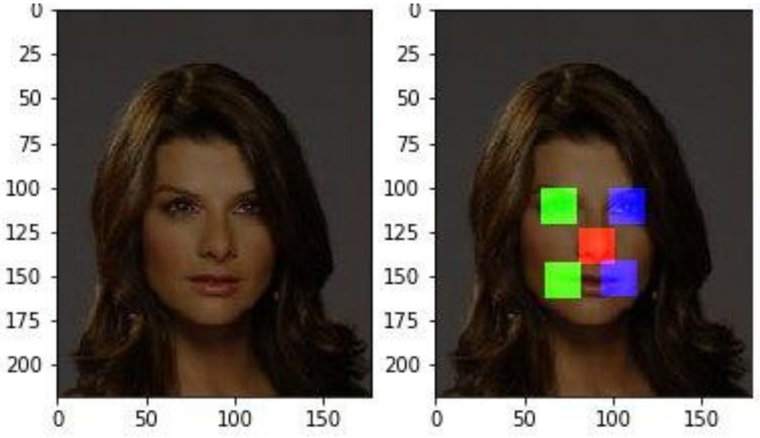
General idea

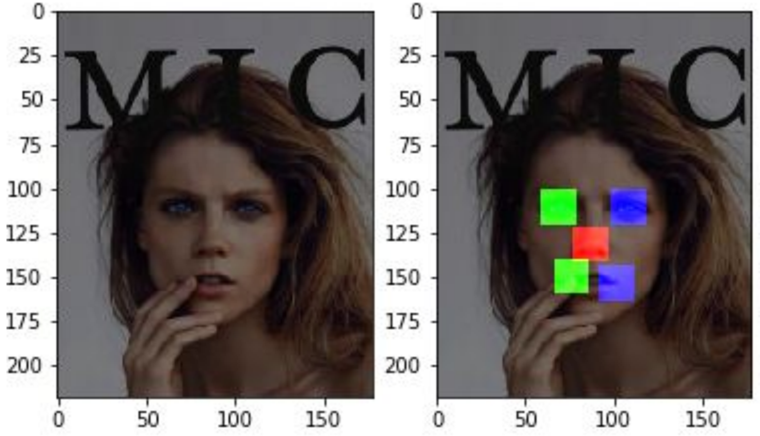


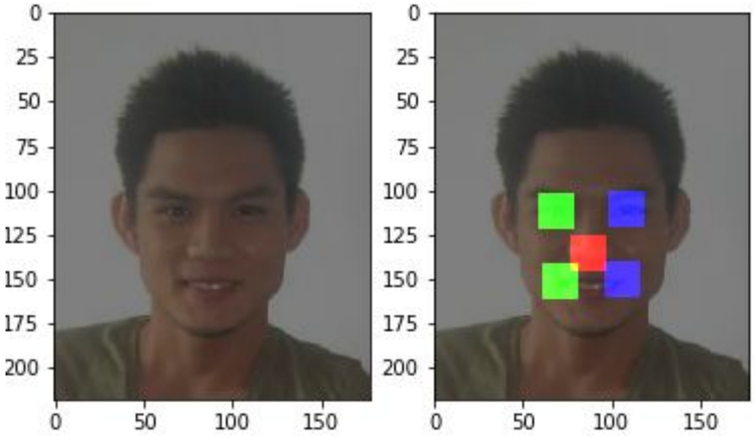
CelebA Dataset

202599 images

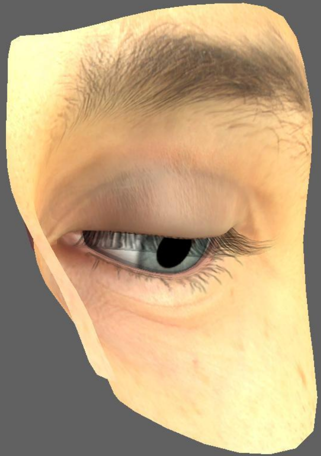
10177 celebrities

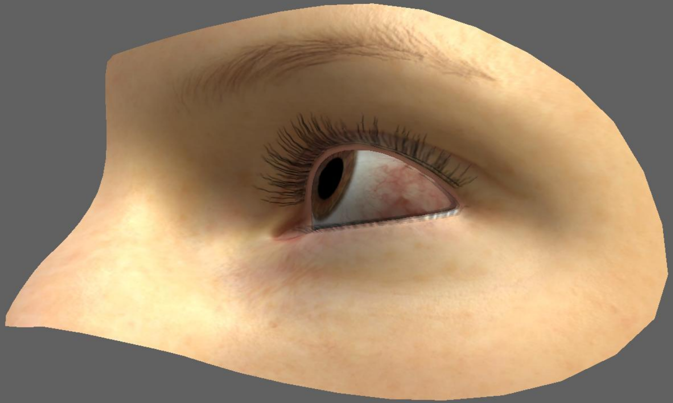






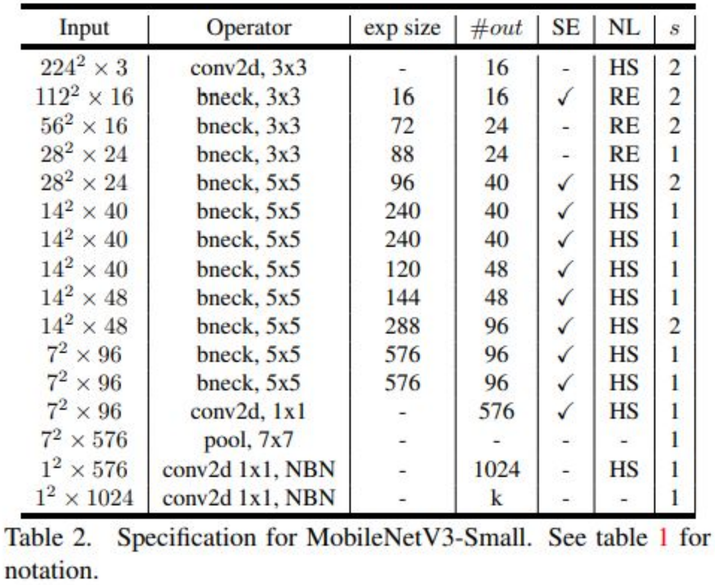
UnityEyes Dataset





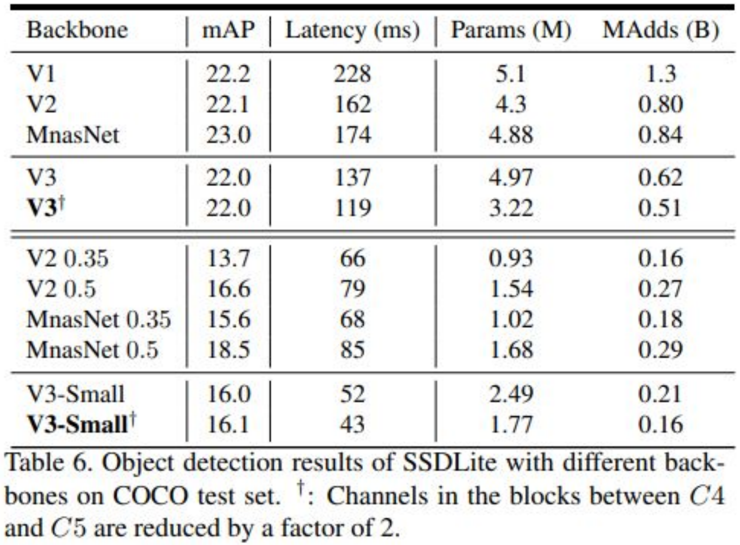
53894 procedurally generated 3D rendered images

MobileNetV3



Input	Operator	exp size	$\#out$	SE	NL	s
$224^2 \times 3$	conv2d, 3x3	-	16	-	HS	2
$112^2 \times 16$	bneck, 3x3	16	16	✓	RE	2
$56^2 \times 16$	bneck, 3x3	72	24	-	RE	2
$28^2 \times 24$	bneck, 3x3	88	24	-	RE	1
$28^2 \times 24$	bneck, 5x5	96	40	✓	HS	2
$14^2 \times 40$	bneck, 5x5	240	40	✓	HS	1
$14^2 \times 40$	bneck, 5x5	240	40	✓	HS	1
$14^2 \times 40$	bneck, 5x5	120	48	✓	HS	1
$14^2 \times 48$	bneck, 5x5	144	48	✓	HS	1
$14^2 \times 48$	bneck, 5x5	288	96	✓	HS	2
$7^2 \times 96$	bneck, 5x5	576	96	✓	HS	1
$7^2 \times 96$	bneck, 5x5	576	96	✓	HS	1
$7^2 \times 96$	conv2d, 1x1	-	576	✓	HS	1
$7^2 \times 576$	pool, 7x7	-	-	-	-	1
$1^2 \times 576$	conv2d 1x1, NBN	-	1024	-	HS	1
$1^2 \times 1024$	conv2d 1x1, NBN	-	k	-	-	1

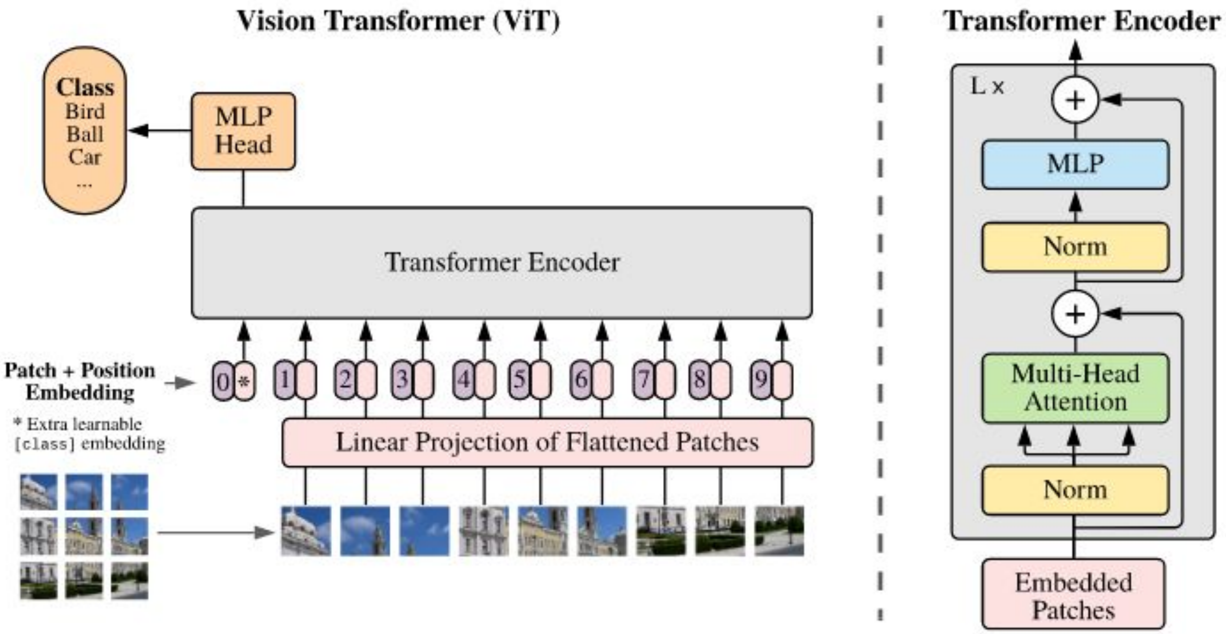
Table 2. Specification for MobileNetV3-Small. See table 1 for notation.



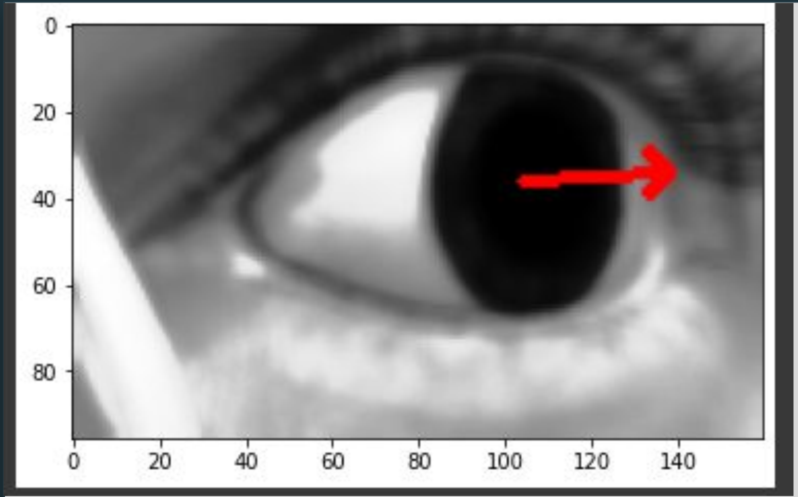
Backbone	mAP	Latency (ms)	Params (M)	MAdds (B)
V1	22.2	228	5.1	1.3
V2	22.1	162	4.3	0.80
MnasNet	23.0	174	4.88	0.84
V3	22.0	137	4.97	0.62
V3[†]	22.0	119	3.22	0.51
V2 0.35	13.7	66	0.93	0.16
V2 0.5	16.6	79	1.54	0.27
MnasNet 0.35	15.6	68	1.02	0.18
MnasNet 0.5	18.5	85	1.68	0.29
V3-Small	16.0	52	2.49	0.21
V3-Small[†]	16.1	43	1.77	0.16

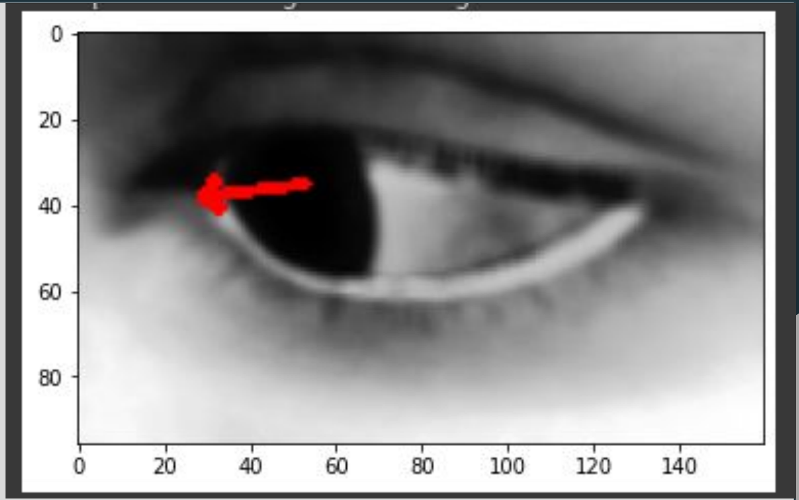
Table 6. Object detection results of SSDLite with different backbones on COCO test set. [†]: Channels in the blocks between $C4$ and $C5$ are reduced by a factor of 2.

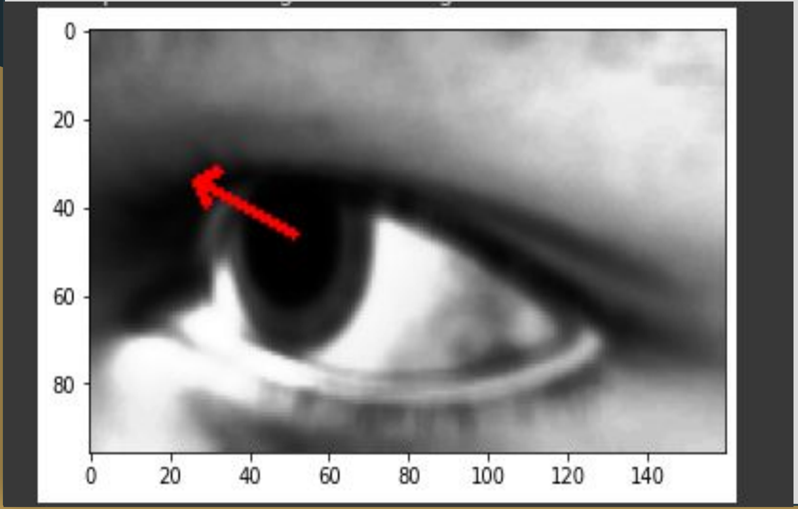
Vision Transformer

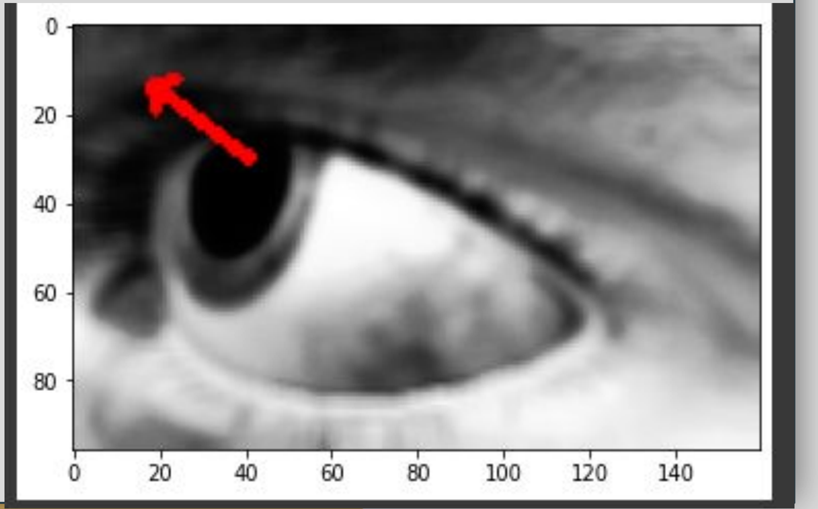


RESULTS

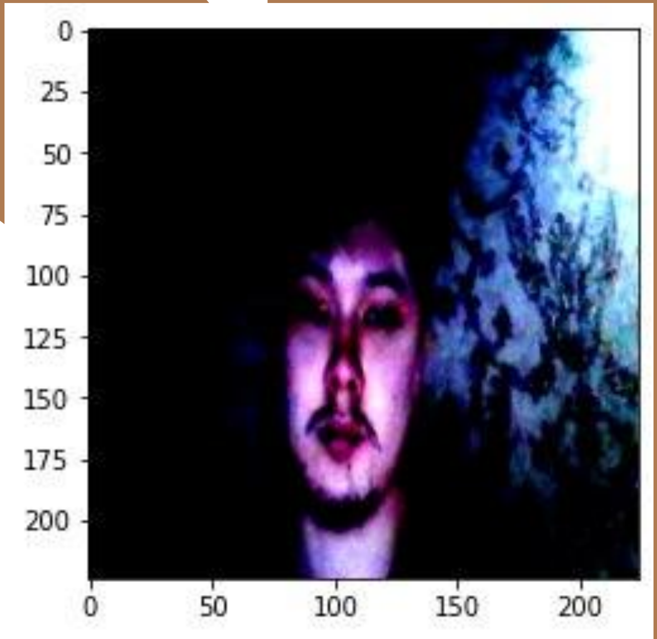

0
20
40
60
80
0
20
40
60
80
100
120
140

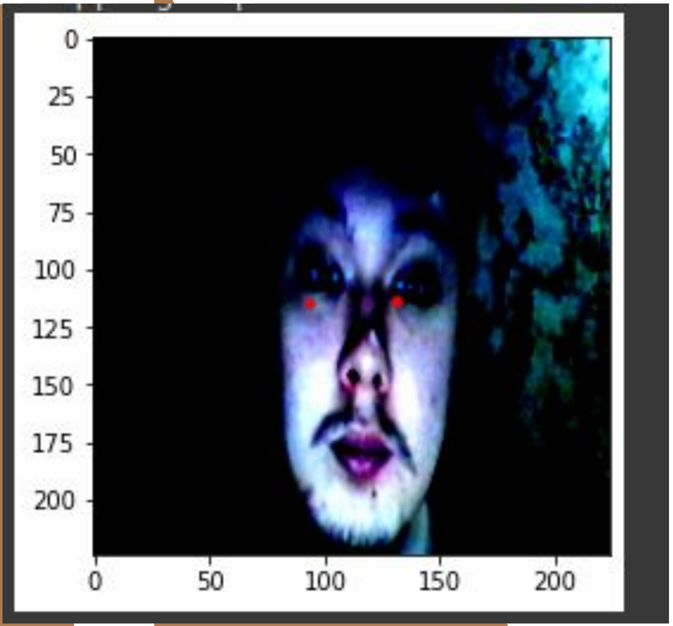

0
20
40
60
80
0
20
40
60
80
100
120
140


0
20
40
60
80
0
20
40
60
80
100
120
140

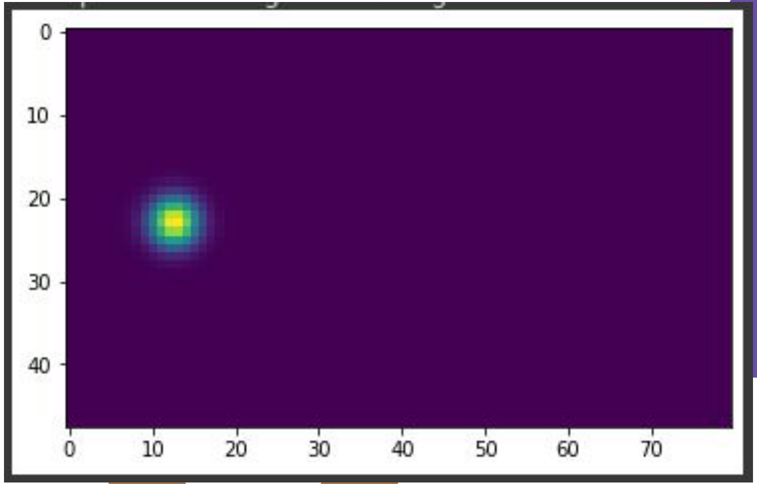

0
20
40
60
80
0
20
40
60
80
100
120
140

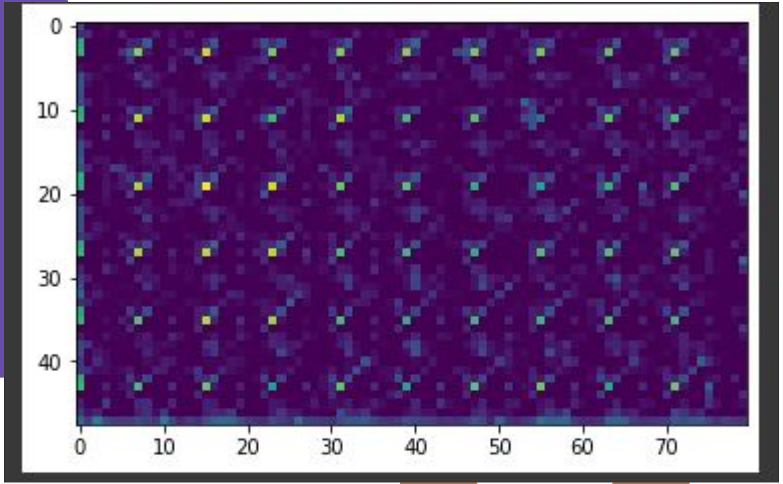
Results


0
25
50
75
100
125
150
175
200
0
50
100
150
200


0
25
50
75
100
125
150
175
200
0
50
100
150
200

Fun Fact





Future Works

Conclusion