STAATLICH ANERKANNTE HOCHSCHULE

Vehicle Pose estimation using Centernet on monocular street images and recommend suitable driving skill

Case Study 2
PPT version 2.0

Presented by:
Manmeet Kumar Chaudhuri (11011780)
Rishabh Garg (11011875)
Rohit Bewoor (11011831)
Sanika Medankar (11011861)
Shekhar Singh (11011694)

SRH HOCHSCHULE HEIDELBERG
Intelligence in Learning

DAS CORE PRINZIP

Content

STAATLICH
ANERKANNTE
HOCHSCHULE

- Introduction
  - Problem Statement and Description of Data provided
- General Background on Image processing and CNNs
  - Choice of Centernet for the use case
- Methodology
  - Understanding how to use the data
  - Preprocessing
  - Model Training
- Results
- Conclusion
- Future Scope
- Demo
- Use Case: Driving Skill Recommended in real world traffic scenario
- Output visualisation after each Epoch

Introduction

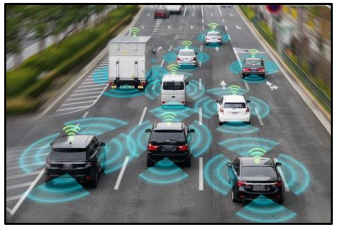
STAATLICH ANERKANNTE HOCHSCHULE

- Autonomous driving is currently a hot field with active ongoing research

- Kaggle competition from Baidu's Autonomous Driving Lab
  - **Problem statement: Estimate the absolute pose of vehicles (6 Degrees of Freedom aka DoF) from a real world traffic environment image**

- Our work:
  - Find a best-in-class detector suitable for use case
  - Perform comparative study of different backbones
  - Use the best model (maximum mAP) to make inference and recommend an appropriate driving skill level based on our own logic

Source: https://www.logigear.com/blog/software-testing/testing-autonomous-vehicles/



Source: https://theconversation.com/what-if-autonomous-vehicles-actually-make-us-more-dependent-on-cars-98498

Data and Expected Output

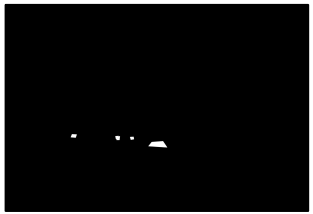
STAATLICH ANERKANNTE HOCHSCHULE

- **Data Provided:**
  - Training CSV file, 4,200 jpeg images (3,384 x 2,710 dimensions)
  - Layout:
    - ImageId,[model yaw pitch roll x y z … once for each car in the image]
    - Example below has two cars with model types 16 and 56.

```
ImageId,PredictionString
ID_8a6e65317,16 0.254839 -2.57534 -3.10256 7.96539 3.20066 11.0225
56 0.181647 -1.46947 -3.12159 9.60332 4.66632 19.339
```

  - Ignore mask (containing vehicles not to be considered) images provided for approx. 80% of the Train and Test images
  - Test Images = 2,000 jpeg images

Train Image



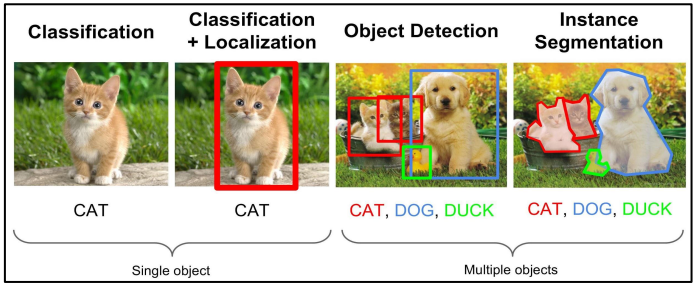
Ignore Mask

- **Expected output:**
  - CSV file containing the predicted 6 DoF for each car detected in image.

General Background

STAATLICH ANERKANNTE HOCHSCHULE

- Classification vs Detection vs Segmentation



- Object Detection using **Monocular image vs LiDAR**
  - Both popular in field today. Waymo using LiDAR and Tesla using Monocular

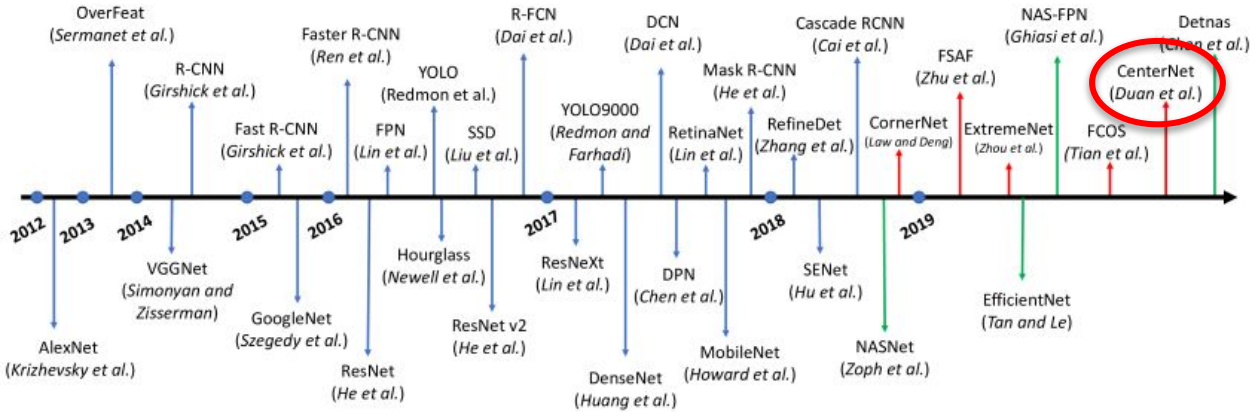
Source: https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852

	Pros	Cons
Monocular	1) Low cost	1) Susceptible to darkness and poor Weather (like fog)
LiDAR	1) Accurate 2) Research ongoing for several years in field	1) Susceptible to interference from other LiDAR and poor weather (like rain)

General Background

STAATLICH ANERKANNTE HOCHSCHULE

- Research over the last decade
  - Two stage detectors: Region based (R-CNN, Fast R-CNN, Faster R-CNN, etc)



Source: https://www.groundai.com/project/recent-advances-in-deep-learning-for-object-detection/1

  - One stage detectors: YOLO, Single Shot Detector, Centernet
  - Deep networks for classification: Resnet, VGG, Alexnet, ResNeXt

Why Centernet?

STAATLICH ANERKANNTE HOCHSCHULE

- Fig. 1: Centernet performs better against one-stage and two-stage image detectors like YOLO, FasterRCNN, etc
- Fig. 2: Centernet performance is comparable to other 3D detectors like Deep3DBox and Mono3D. ***However, CenterNet is two order of magnitude faster than these.***
- Fig. 3: Centernet with Hourglass performs better in compare to other backbones i.e. DLA, ResNet. ResNet backbone is approximately 10 times faster than Hourglass and c. 2.5 times faster than DLA backbone
- **This Use Case: Centernet with ResNet backbone is used due to suitably high accuracy and speed**

Speed-accuracy trade-off on COCO dataset against various real-time detectors

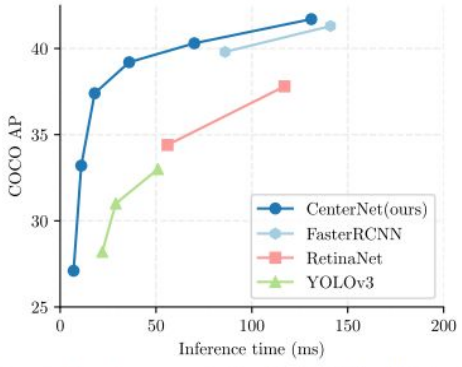


Fig. 1

3D object detection comparison with other popular detectors

	AP		
	Easy	Mode	Hard
Deep3DBox [38]	98.8	97.2	81.2
Ours	90.2±1.2	80.4±1.4	71.1±1.6
Mono3D [9]	95.8	90.0	80.6
Ours	97.1±0.3	87.9±0.1	79.3±0.1

Fig. 2

Speed Accuracy Tradeoff on COCO dataset with different backbones

	AP			FPS		
	N.A.	F	MS	N.A.	F	MS
Hourglass-104	**40.3**	**42.2**	**45.1**	14	7.8	1.4
DLA-34	37.4	39.2	41.7	52	28	4
ResNet-101	34.6	36.2	39.3	45	25	4
ResNet-18	28.1	30.0	33.2	**142**	**71**	**12**

Fig. 3

All Sources: Centernet Paper: Objects as Points

Centernet Detector - Overview

STAATLICH ANERKANNTE HOCHSCHULE

- It models objects as single point which represents its center. Based on Cornernet and uses center-pooling.

- One can regress any required object properties (KEYPOINTS)
  - object categories for classification task
  - 17 points for human pose detection
  - size, location, pose, etc
- Paper used various backbones: Resnet-18, Resnet-101, Hourglass-104, DLA-34

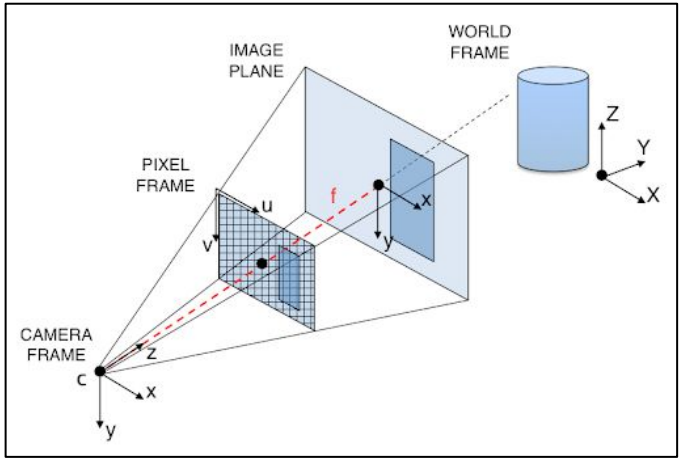
- **This Use-case:**
  - **Backbones: Resnet-18 and ResNeXt-50_32**
  - **Regressing on 7 values to infer pose (6 DoF)**

Basics of Image Transformation

STAATLICH ANERKANNTE HOCHSCHULE

- **World Frame (Ow)**: a fixed coordinate system to represent objects in world
  - model vertices json file
- **Camera Frame (Oc)**: coordinate system with camera at origin
  - Pose information (yaw,pitch,roll,x,y,z)
- **Image Frame**: Pixel coordinates (after projecting the Camera Frame objects to image plane)
- **Camera Intrinsic Properties (k)**: Parameters internal to the camera/ digitization setup
- Equations to transform:
  - Oi = k [R | t] Ow ⇒ Maps to x,y,z
  - Divide by z to get the true pixel coordinates in image

Frame changes from World => Camera => Image



Source: http://people.bu.edu/chenyua/CS585finalproject/CS585_Project_Report.html

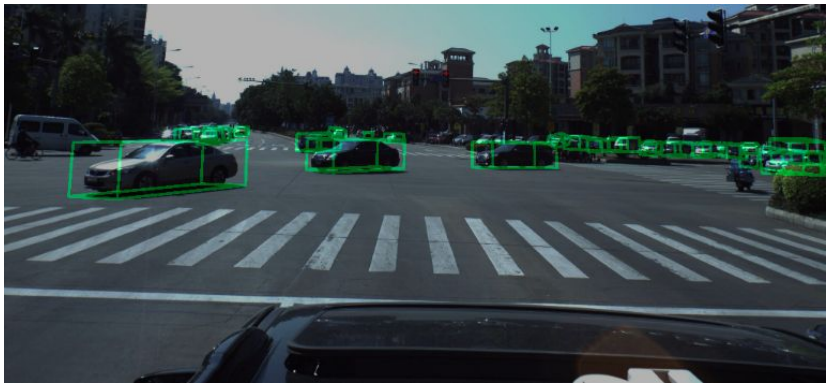
Visualising in Camera Coordinates

STAATLICH
ANERKANNTE
HOCHSCHULE

- Projecting a bounding box (created from the model.json files) from World coordinates to Pixel coordinates.

Original Image

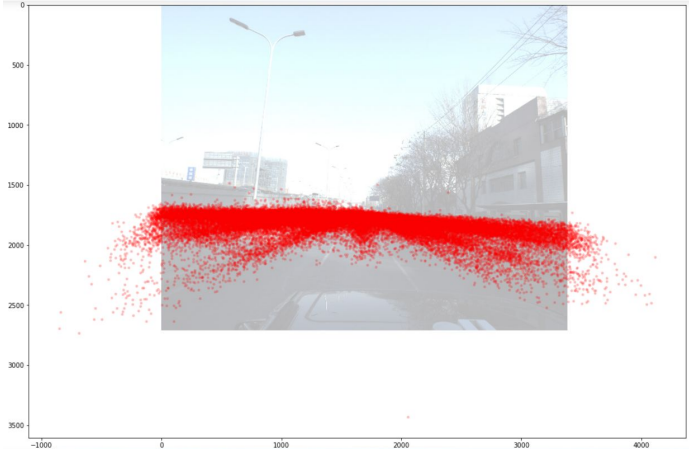
Bounding Box projected onto Original Image



Preprocessing

STAATLICH ANERKANNTE HOCHSCHULE

- Removed bad images: Corrupted images or overexposed images
- Divided image in half vertically and retained only bottom part (cars only on level ground only)
- Around 2% of the ground truth prediction string data, mapping to image pixel coordinates was to a point outside the image, hence resized the image with a margin of (width/4) on the sides
- Values to be regressed are converted/transformed into new features to allow better granularity in predictions



Plot containing all the vehicles on an image

S. No.	Feature		
	Original	Remarks	Transformed
1	x	Along the width of the road	x/100
2	y	Perpendicular to the road	y/100
3	z	Along the length of the road	ln(z/100)
4	yaw	Rotation along the x-axis	unchanged
5	pitch	Rotation along the y-axis	sin(pitch)
			cos(pitch)
6	roll	Rotation along the z-axis	roll + Pi

Transforming the 6 DoFs

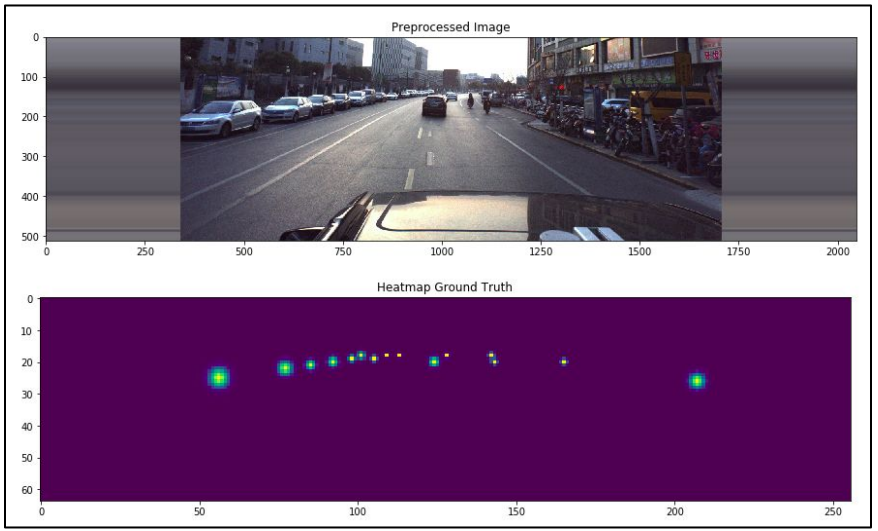
Preprocessing

STAATLICH ANERKANNTE HOCHSCHULE

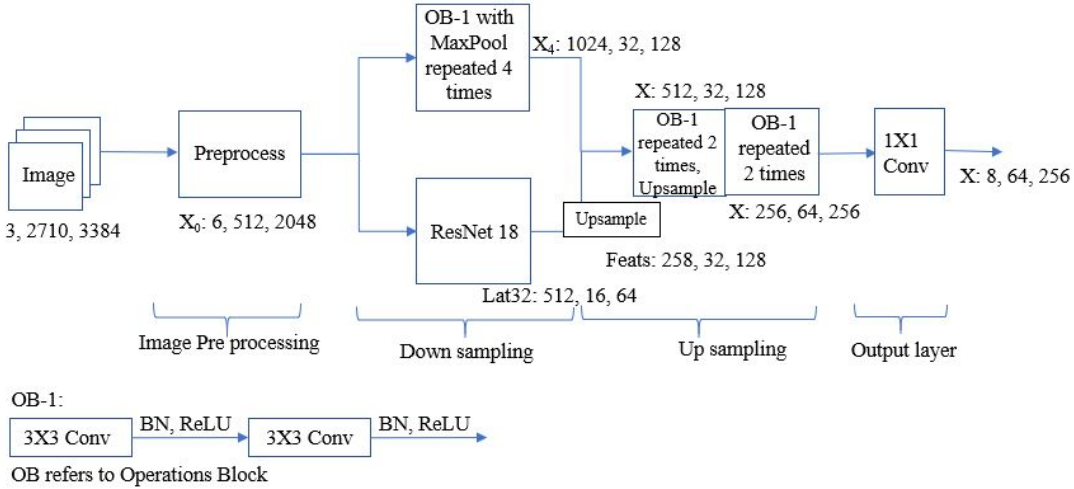
- Original input image dimensions = 2710 x 3384 x 3
- Preprocessed image dimension = 512 x 2048 x 6 (3 for RGB + 1 for Ignore Mask + 2 for Mesh)
- Heatmap (containing vehicles as center points) dimension = 64 x 256

Original Image

Image after Preprocessing and the Ground Truth Heatmap



CenterNet High Level Block Diagram



- Output: (8, 64, 256) containing 8 planes
  - Plane 1 is for the prediction centerpoints/heatmap
  - Rest 7 planes are for the values to be regressed (which allow 6 DoF pose estimation later on)

Model Hyperparameters

STAATLICH ANERKANNTE HOCHSCHULE

- Optimizer:
  - AdamW: It is a variant of Adam with Decoupled Weight Decay
- Learning rate: 0.0009
- Loss Function: Comprises of weighted Mask loss and Regr loss
  - Mask loss (corresponding to the predicted and actual center point/heatmap): Weight = 0.5
    - Binary Cross Entropy loss, or $-\sum_{i=1}^{n} \hat{y}_i log y_i + (1-\hat{y}_i) log(1-\hat{y}_i)$
    - Focal Loss $$\frac{-1}{N}\sum_{xyc}\begin{cases}(1-\hat{Y}_{xyc})^{\alpha}\log(\hat{Y}_{xyc}) & \text{if } Y_{xyc}=1\\ (1-Y_{xyc})^{\beta}(\hat{Y}_{xyc})^{\alpha} & \text{otherwise}\\ \log(1-\hat{Y}_{xyc}) & \end{cases}$$

    alpha = 2; beta = 4 for CenterNet
  - Regr loss (corresponding to the predicted and actual 7 planes that are proxies for 6 DoFs): Weight = 1.0
    - L1 Loss $\sum_{i=1}^{n} |y_{true} - y_{predicted}|$

Evaluation Matrix

STAATLICH ANERKANNTE HOCHSCHULE

- Metric used to compare the accuracy is mean average precision (mAP)
- Measures affecting the mAP:
  - Translational difference: Euclidean distance
  - Rotational difference: Quaternion rotation
- Using the above two differences, Precision and Recall are calculated
- Average precision (AP) is calculated based on the precision and recall values

$$AP = \frac{1}{11} \sum_{\text{Recall}_i} \text{Precision}(\text{Recall}_i)$$

- mAP is calculated by taking the mean of AP over all the 6 DoF for only the translational and rotational thresholds

Result

STAATLICH ANERKANNTE HOCHSCHULE

Comparative Study of different backbones

- For given epochs, using Focal loss improves mAP by about 0.02
- mAP rises with more training
- Average inference time / image for all models is consistently around 2-3 seconds

Centernet models – Time and Precision metrics

SN	Backbone	Epochs Trained	Loss Function	Mask Images Used	Time (seconds)		mAP
					Total	Average Per Image	
1	Resnet-18	10	Focal	Yes	1224	2.87	0.10616
2	Resnet-18	6	Focal	Yes	786	2.45	0.07877
3	Resnet-18	4	Focal	Yes	866	2.03	0.07456
4	Resnet-18	4	BCE	Yes	1264	2.96	0.04754
5	Resnext-50_32	4	Focal	Yes	941	2.21	0.08551
6	Resnext-50_32	4	BCE	Yes	965	2.26	0.06229

- As ResNeXt-50_32 has a larger depth and width, one would expect the inference time to be slightly higher compared to that of Resnet-18. However, this is not the case for ResNeXt-50-4epoch-BCE vs ResNet-18-4epoch-BCE with 2.26 sec vs 2.96
- The best model: Resnet-18 backbone, for 10 epochs and with Focal loss function

Result

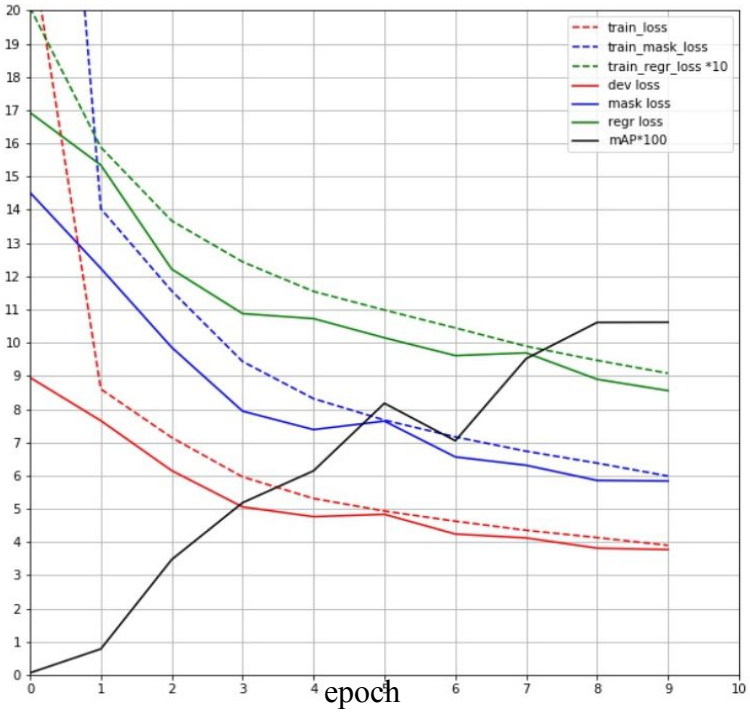
STAATLICH ANERKANNTE HOCHSCHULE

Best Model training statistics

The following observations can be made from the graph:

- The training loss and development losses fall steeply during the initial 3 epochs
- After 3 epochs, the training and development losses fall gradually
- mAP rises with each epoch. There is a fall in mAP from 5th epoch to 6th epoch
- Overall, the trend of losses and the mAP with each epoch is as expected

Training statistics against Epoch



Conclusion

STAATLICH
ANERKANNTE
HOCHSCHULE

Based on our experimenting with different models and their training hyperparameters, we conclude the following:

- The ResNeXt-50_32 model outperforms a Resnet-18 in terms of accuracy

- If we were able to train for say 50-100 epochs, the mAP most probably would be much higher than that achieved with 10 epochs.

- We started off with very simple models with following parameters:
  - simple BCE loss function,
  - no use of training mask images,
  - no pre-checks to ensure only sanitized training data and
  - no transformation on the variables to be predicted (pose information).

  The mAP did not even cross very single digits.

- Only after incorporating training data sanity checks, using a better loss function (Focal loss) and figuring out how to use the mask images, were we able to improve the accuracy to around 10%.

Future Scope

STAATLICH
ANERKANNTE
HOCHSCHULE

Possible enhancements that could be made in our approach are as follows:

- Data augmentation like h-flip, 3 axis rotate, color, noise, blur, etc. could be incorporated in the dataset
- A stacked Hourglass backbone can also be used to check the accuracy levels
- This experiment is done on the ApolloCar3D dataset. KITTI dataset could be used to improve model accuracy via transfer learning IF they provide the ground truth values for pose information.
- A further study could look at how each of the 6 DoF variables contributes towards the overall error

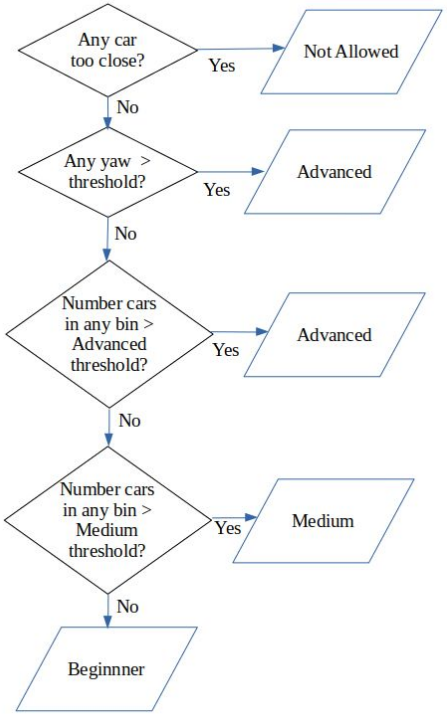
STAATLICH
ANERKANNTE
HOCHSCHULE

Demo

Driving Skills

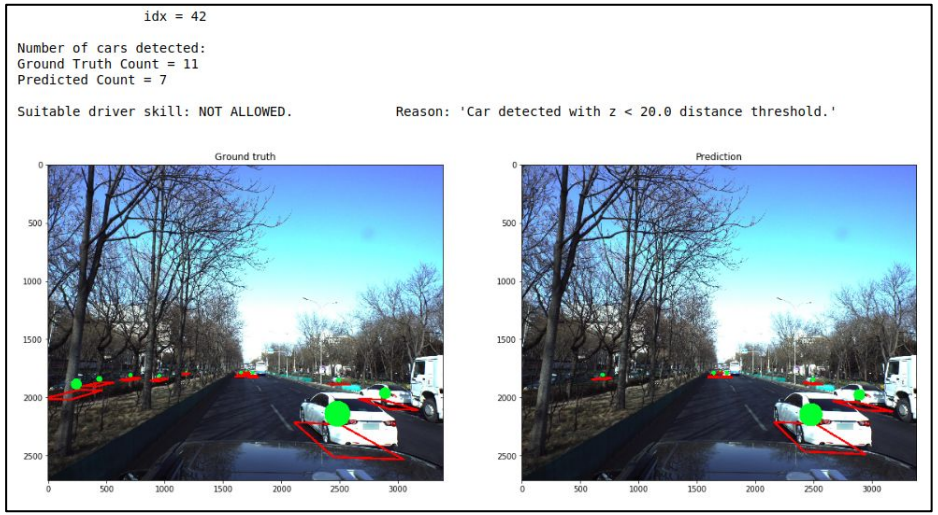
STAATLICH ANERKANNTE HOCHSCHULE

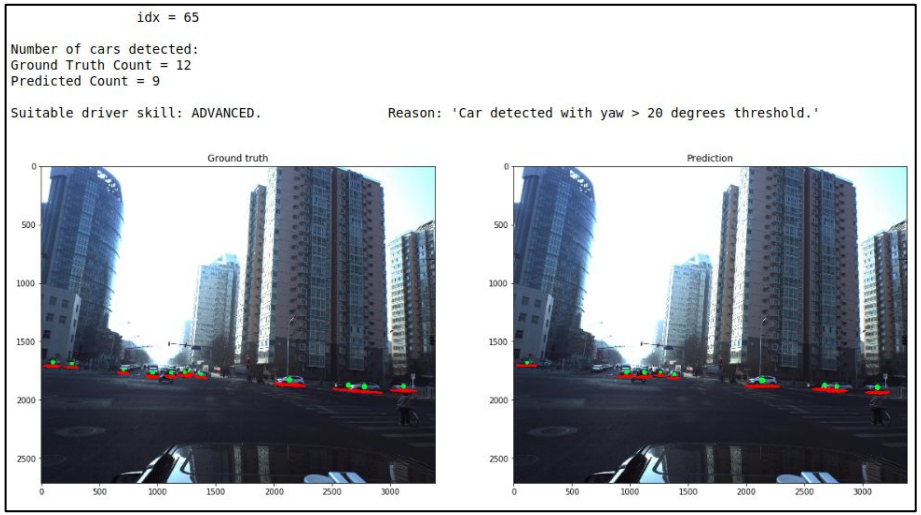
- **Scenario in the future (20 years from now?) :**
  - Autonomous driving is default choice - no one drives manually
  - Manual control is allowed only for person “ADVANCED profile”
  - AI evaluates environment and decides whether safe to hand over control
- Rules as per flowchart:
  - Car too close threshold = 20 units (z)
  - Yaw threshold = 20 degrees (pitch in model)
  - Bin sizes = 20 units
  - Max limit cars / Bin
    - Advanced threshold = 8+ cars detected in any bin
    - Medium threshold = 5 to 7 cards detected in any bin



Driving Skills

STAATLICH
ANERKANNTE
HOCHSCHULE





Case 1: NOT ALLOWED

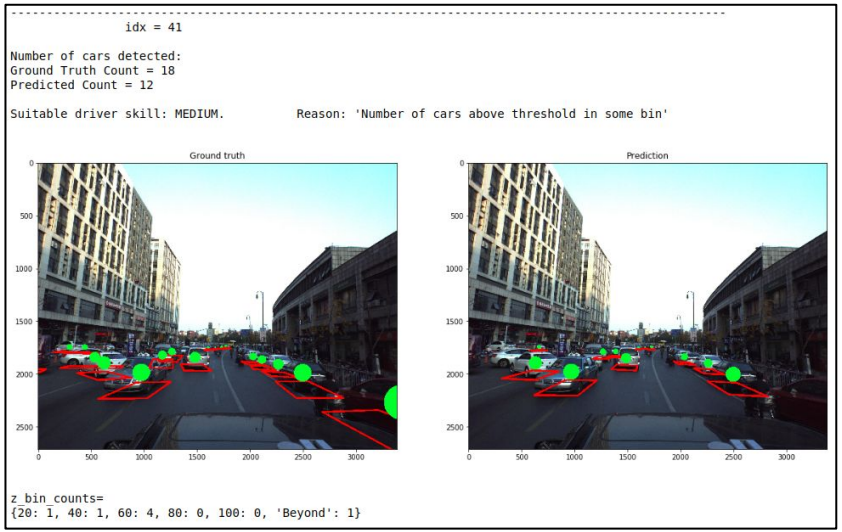
- some car detected is too close (z < 20 units)

Case 2: ADVANCED

- yaw of some car > threshold (i.e. pitch > 20 degrees)

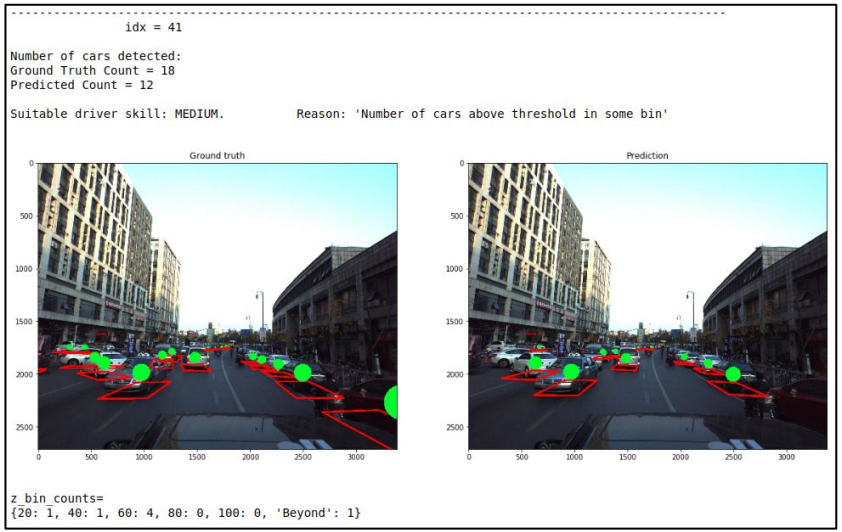
Driving Skills

STAATLICH ANERKANNTE HOCHSCHULE



Case 2: MEDIUM

- no car is too close (all z > 20 units)
- no yaw > threshold (i.e all pitch < 20 degrees)
- max(cars in each bin) < 8



Case 2: BEGINNER

- no car is too close (all z > 20 units)
- no yaw > threshold (i.e all pitch < 20 degrees)
- max(cars in each bin) < 5

STAATLICH
ANERKANNTE
HOCHSCHULE

Epochwise Error rates and the Output plane visualization for some random input from dev set.

Added after presentation on 12.03.2020 based on questions by Prof. Prabhune.

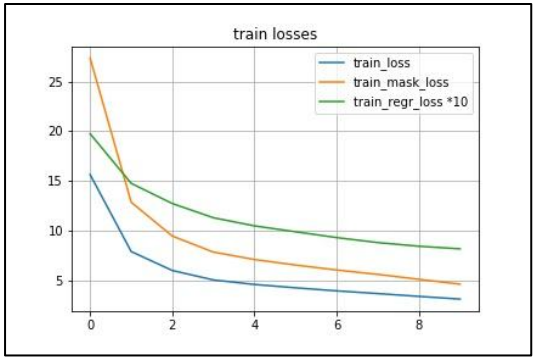
1) First is the error for each epoch - tabular values shown
   a) Earlier slide had one combined graph which was not clear to audience

2) Captured the outputs for the 8 planes (1 of heatmap and 7 for the regression values).
   a) This is for some one input Dev set image (index 23, first position)
   b) Regression values: pitch_cos, pitch_sin, roll, udiff, vdiff, yaw, z

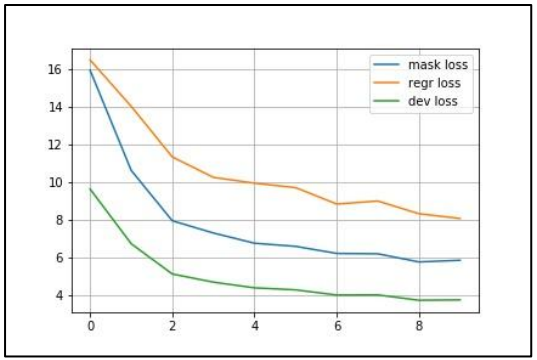
Epochwise Error Rates for one model

STAATLICH ANERKANNTE HOCHSCHULE

Train Loss over 10 epochs



Dev Loss over 10 epochs



Epoch	TRAIN Loss (Total = 50% Mask + 100% Regr)			DEV Loss (Total = 50% Mask + 100% Regr)		
	Mask loss	Regr Loss	Total Loss	Mask loss	Regr Loss	Total Loss
0	27.2	1.98	15.58	15.94	1.65	9.62
1	12.87	1.48	7.92	10.61	1.4	6.7
2	9.47	1.27	6.01	7.93	1.13	5.1
3	7.86	1.14	5.07	7.28	1.02	4.66
4	6.84	1.05	4.47	6.74	0.99	4.36
5	6.81	0.99	4.4	6.57	0.97	4.26
6	6.01	0.94	3.95	6.19	0.88	3.98
7	5.74	0.89	3.76	6.17	0.9	3.98
8	5.09	0.83	3.38	5.74	0.83	3.7
9	4.82	0.81	3.22	5.83	0.8	3.72

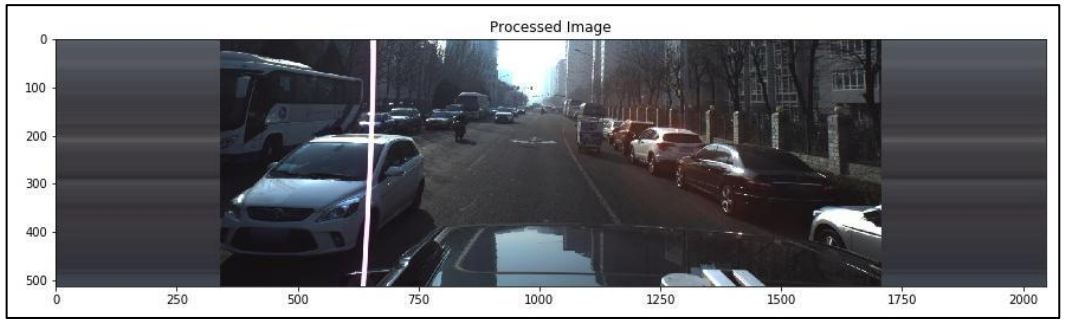
- Resnet-18, Epochs=10, Focal loss
  - TRAIN Loss = 50% of Mask Loss + 100% of Regr Loss
  - Same for DEV Loss
  - Regr losses increased by factor of 10 for scale fitting

Epochwise Output visualisation

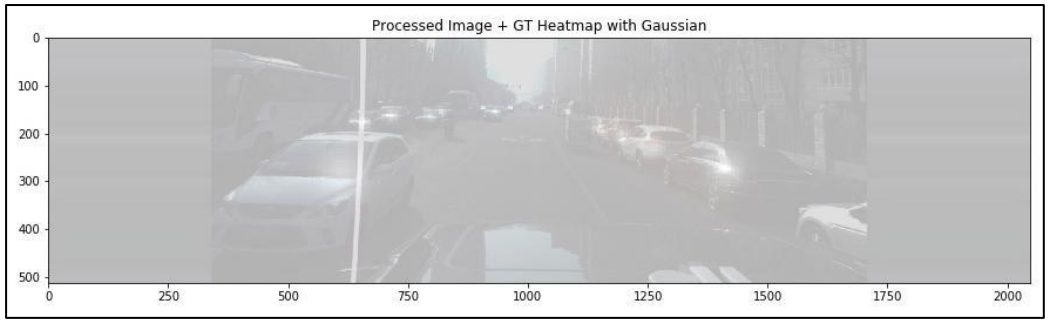
STAATLICH
ANERKANNTE
HOCHSCHULE

Original Image

First image in Dev dataset - index 23



Preprocessed Image

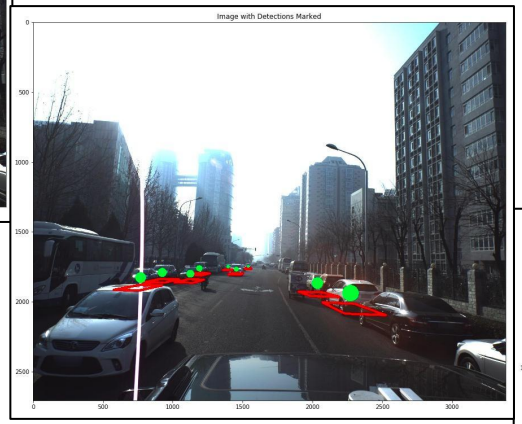


Preprocessed Image with Ground Truth Heatmap Gaussian superimposed

Epochwise Output visualisation

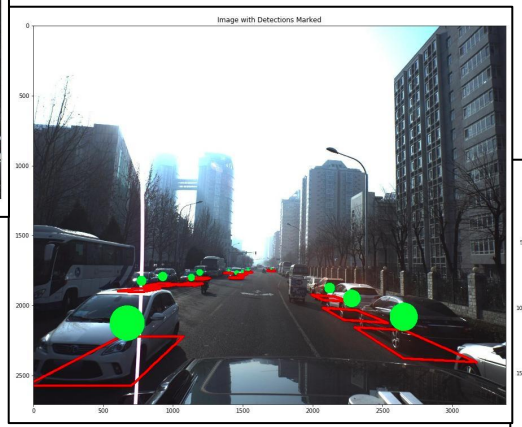
STAATLICH ANERKANNTE HOCHSCHULE

Epoch 0 output

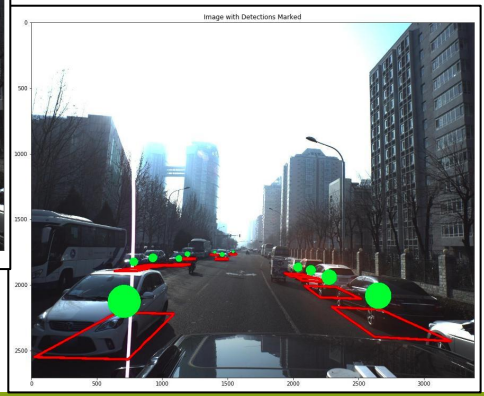


Epoch 2 output

Epoch 5 output



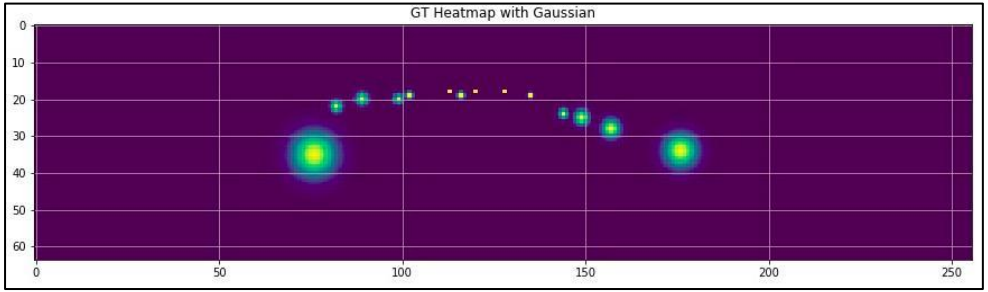
Epoch 9 output



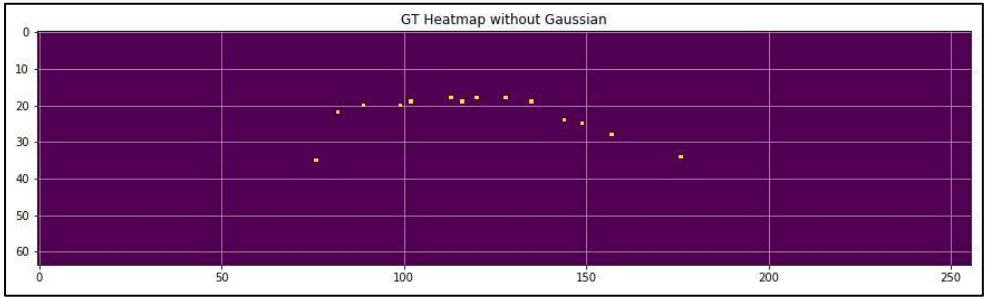
**Note : Total 10 Epochs have been run, but output of only 4 Epochs shown here.

Epochwise Output visualisation - Heatmap

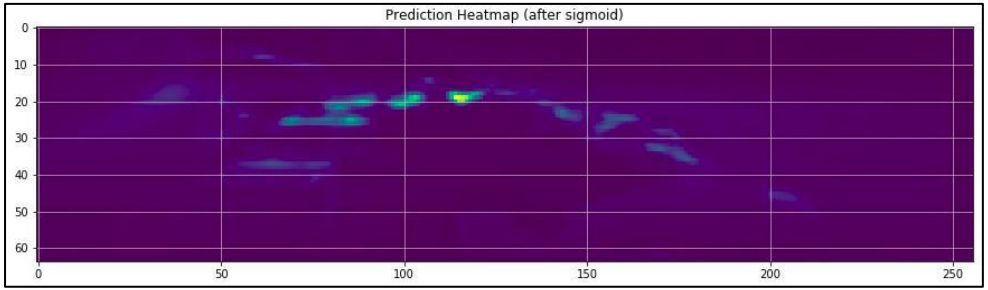
STAATLICH ANERKANNTE HOCHSCHULE



Ground Truth Heatmap Gaussian



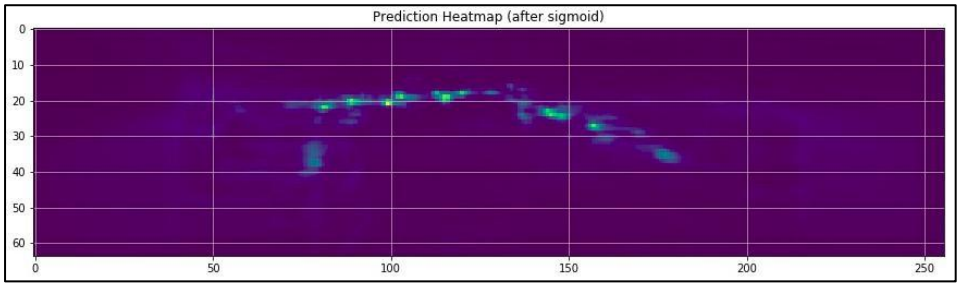
Ground Truth Heatmap non-Gaussian



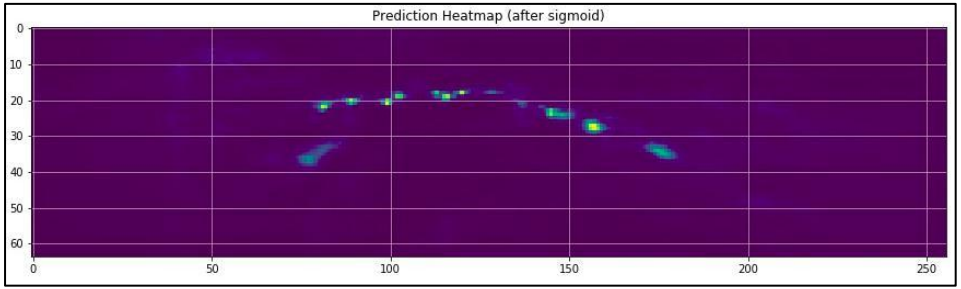
Prediction Heatmap - after Epoch 0

Epochwise Output visualisation - Heatmap

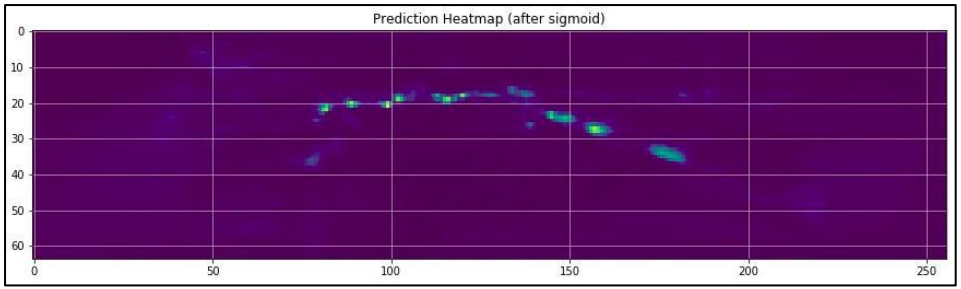
STAATLICH ANERKANNTE HOCHSCHULE



Prediction Heatmap - after Epoch 1



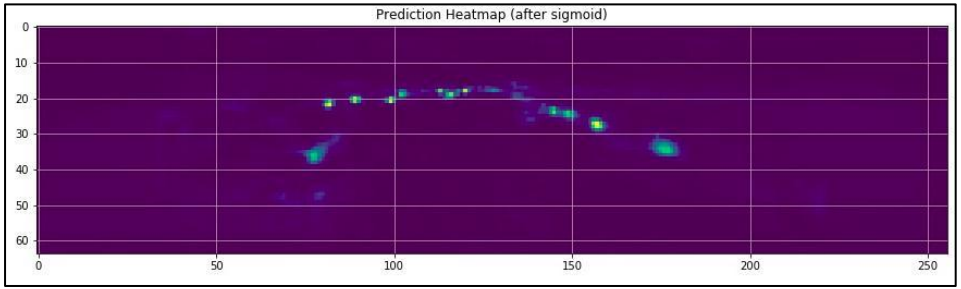
Prediction Heatmap - after Epoch 2



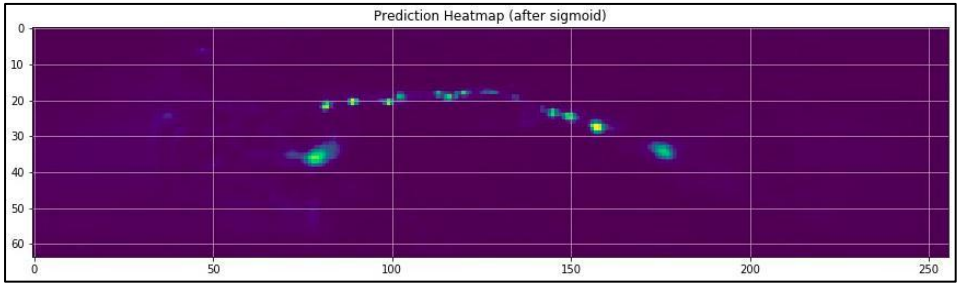
Prediction Heatmap - after Epoch 3

Epochwise Output visualisation - Heatmap

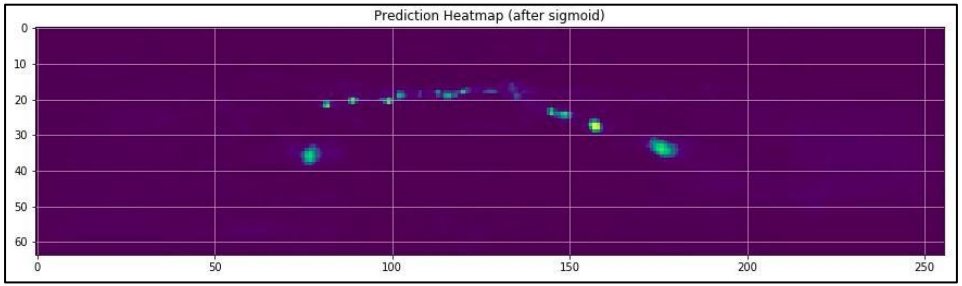
STAATLICH ANERKANNTE HOCHSCHULE



Prediction Heatmap - after Epoch 4



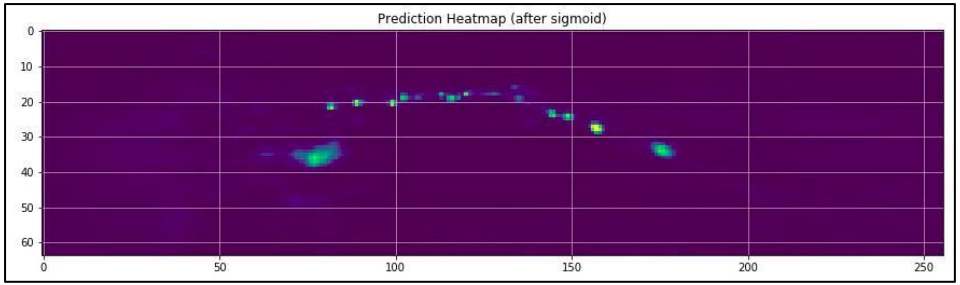
Prediction Heatmap - after Epoch 5



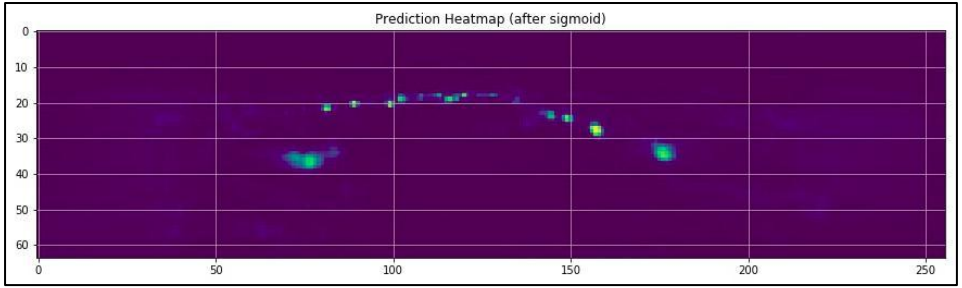
Prediction Heatmap - after Epoch 6

Epochwise Output visualisation - Heatmap

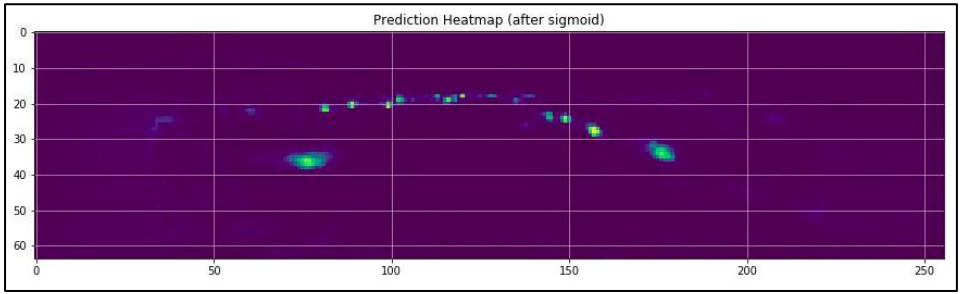
STAATLICH ANERKANNTE HOCHSCHULE



Prediction Heatmap - after Epoch 7



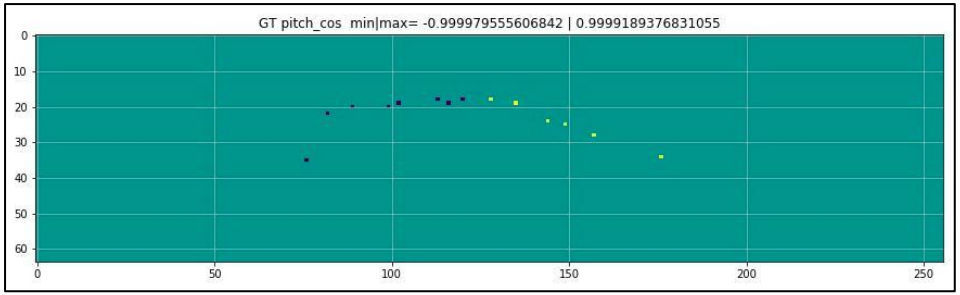
Prediction Heatmap - after Epoch 8



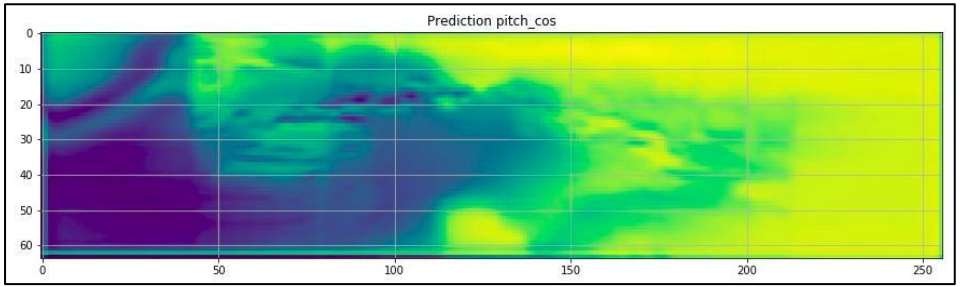
Prediction Heatmap - after Epoch 9

Epochwise Output visualisation - pitch_cos

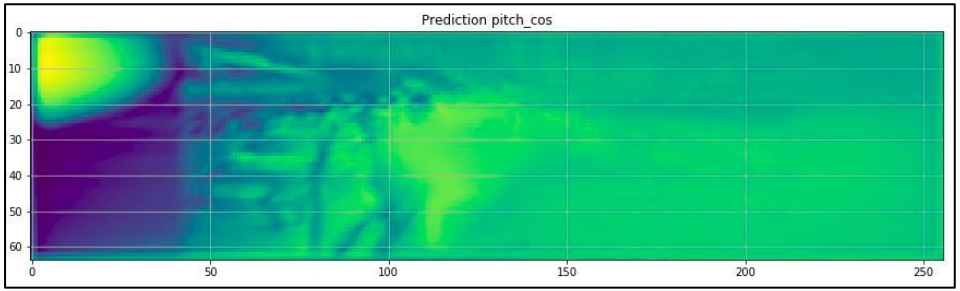
STAATLICH ANERKANNTE HOCHSCHULE



Ground Truth pitch_cos



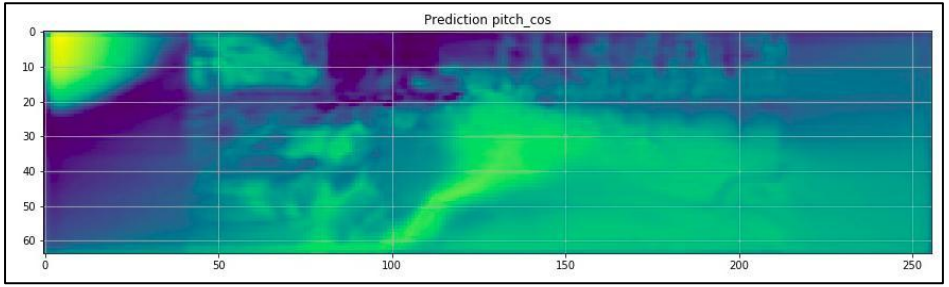
Prediction pitch_cos - after Epoch 0



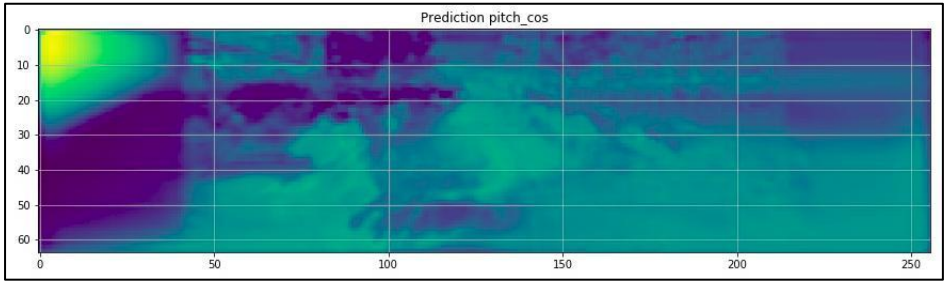
Prediction pitch_cos - after Epoch 1

Epochwise Output visualisation - pitch_cos

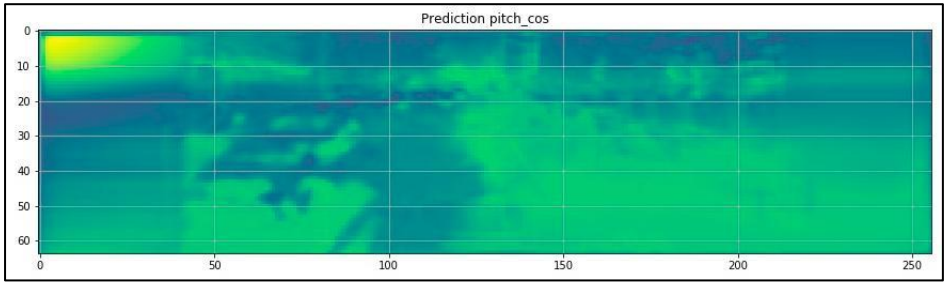
STAATLICH ANERKANNTE HOCHSCHULE



Prediction pitch_cos - after Epoch 2



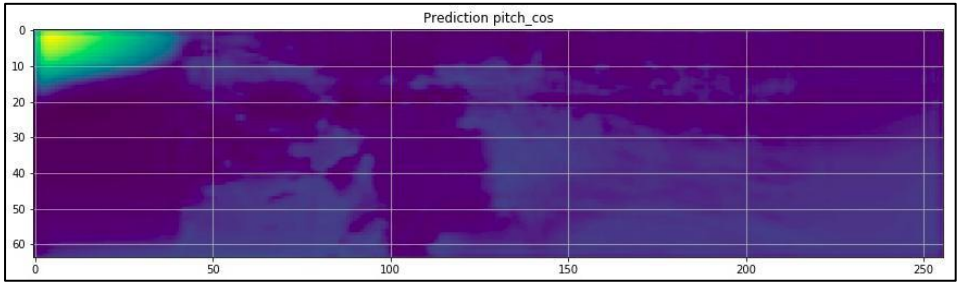
Prediction pitch_cos - after Epoch 3



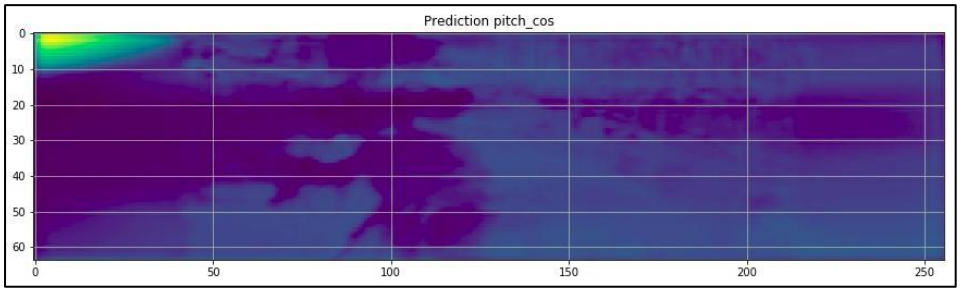
Prediction pitch_cos - after Epoch 4

Epochwise Output visualisation - pitch_cos

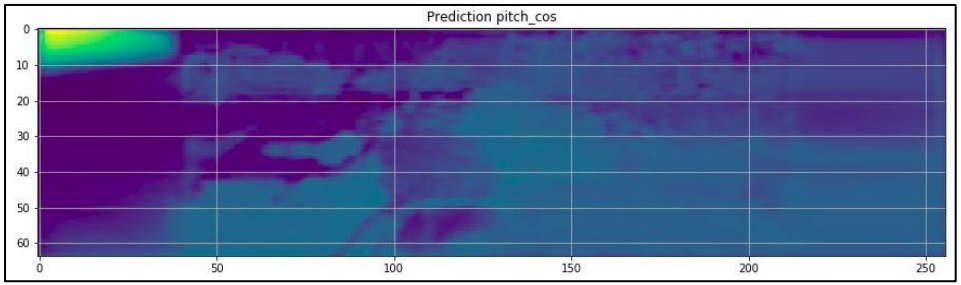
STAATLICH ANERKANNTE HOCHSCHULE



Prediction pitch_cos - after Epoch 5



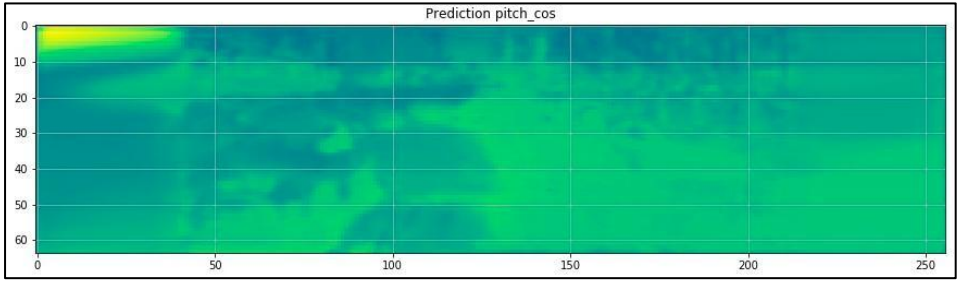
Prediction pitch_cos - after Epoch 6



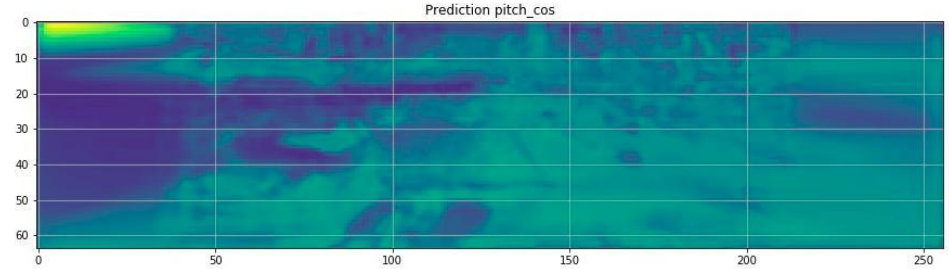
Prediction pitch_cos - after Epoch 7

Epochwise Output visualisation - pitch_cos

STAATLICH
ANERKANNTE
HOCHSCHULE



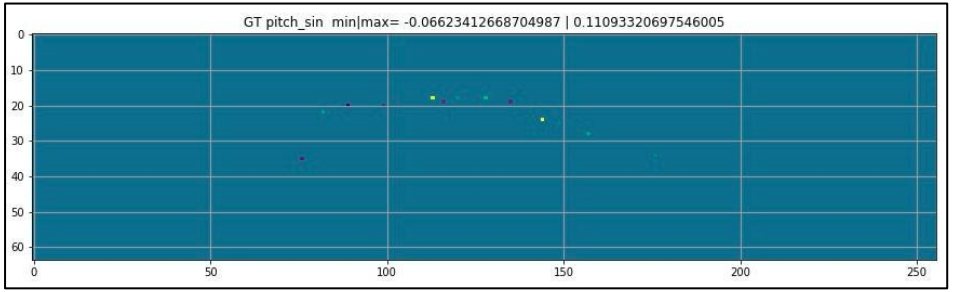
Prediction pitch_cos - after Epoch 8



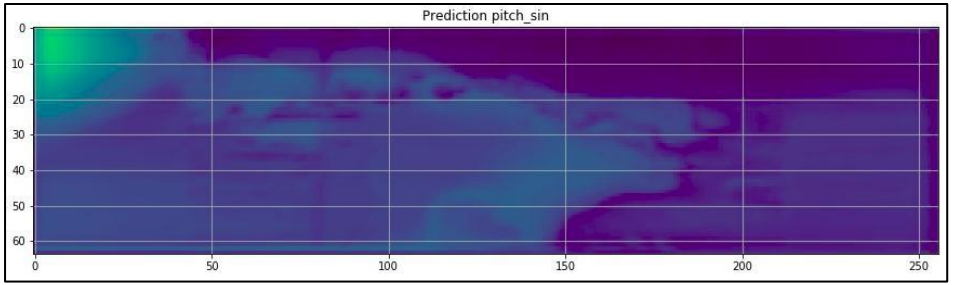
Prediction pitch_cos - after Epoch 9

Epochwise Output visualisation - pitch_sin

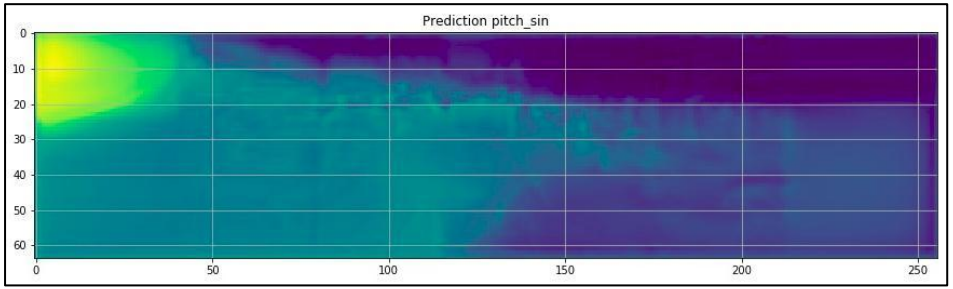
STAATLICH ANERKANNTE HOCHSCHULE



Ground Truth pitch_sin



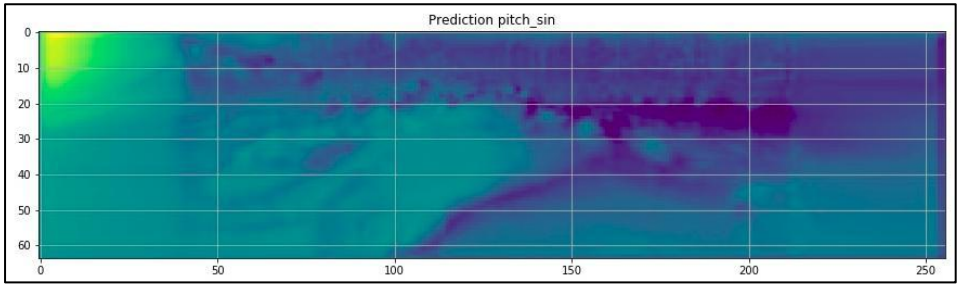
Prediction pitch_sin - after Epoch 0



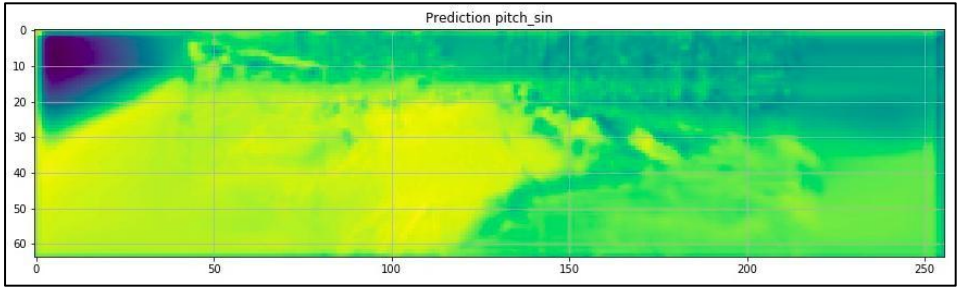
Prediction pitch_sin - after Epoch 1

Epochwise Output visualisation - pitch_sin

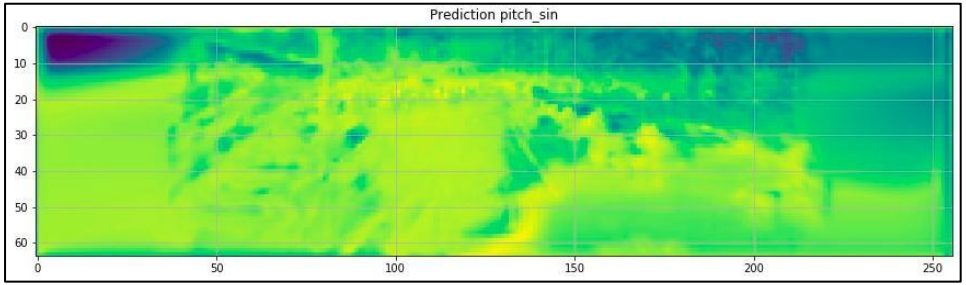
STAATLICH ANERKANNTE HOCHSCHULE



Prediction pitch_sin - after Epoch 2



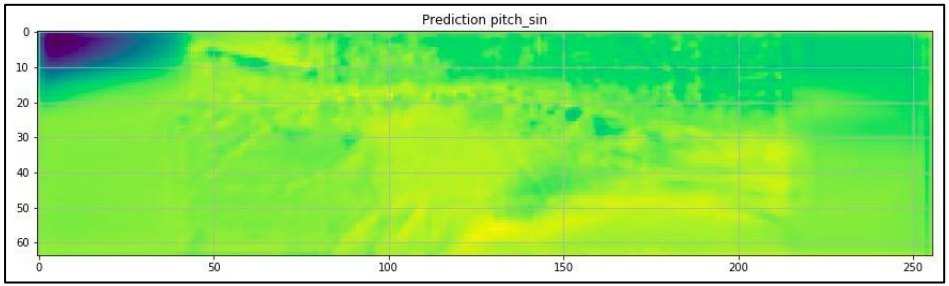
Prediction pitch_sin - after Epoch 3



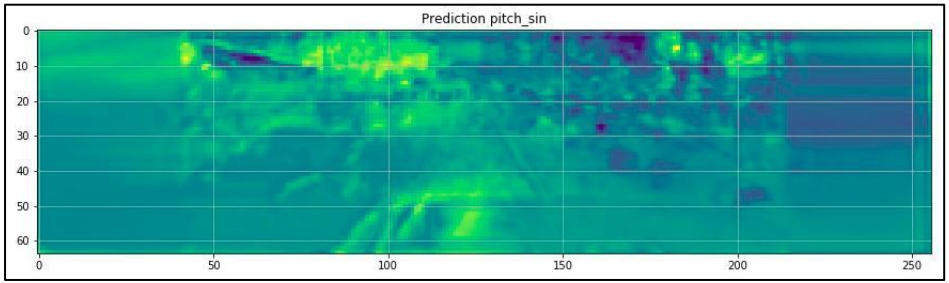
Prediction pitch_sin - after Epoch 4

Epochwise Output visualisation - pitch_sin

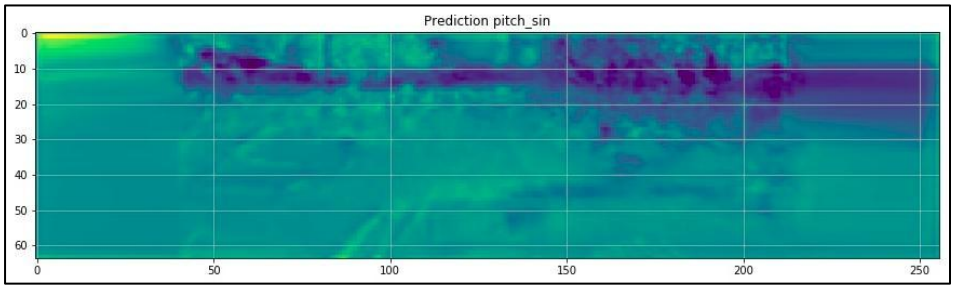
STAATLICH ANERKANNTE HOCHSCHULE



Prediction pitch_sin - after Epoch 5



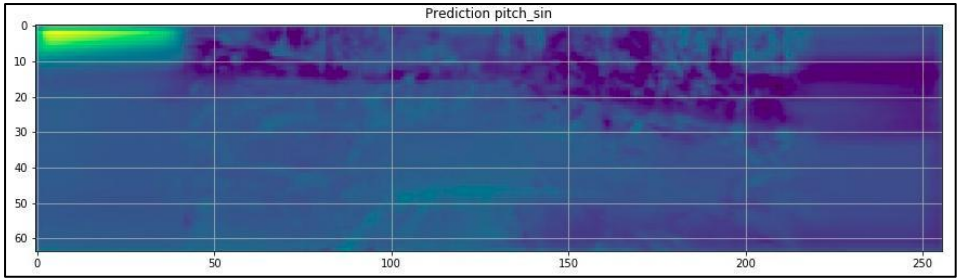
Prediction pitch_sin - after Epoch 6



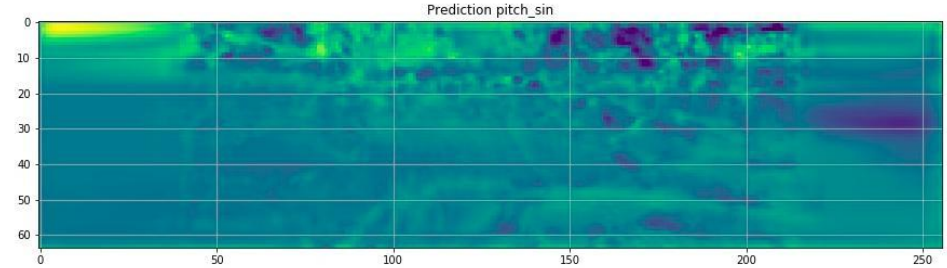
Prediction pitch_sin - after Epoch 7

Epochwise Output visualisation - pitch_sin

STAATLICH
ANERKANNTE
HOCHSCHULE



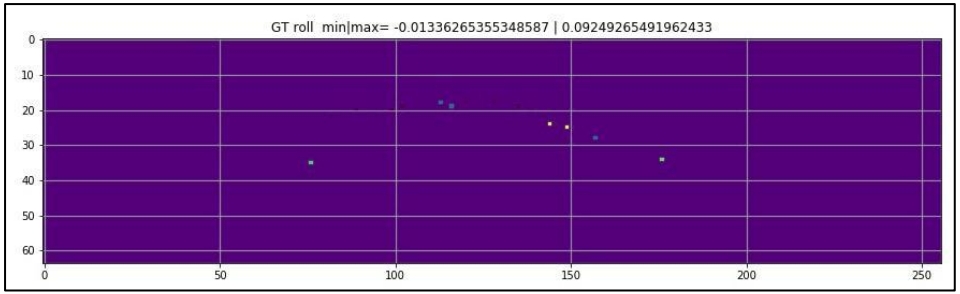
Prediction pitch_sin - after Epoch 8



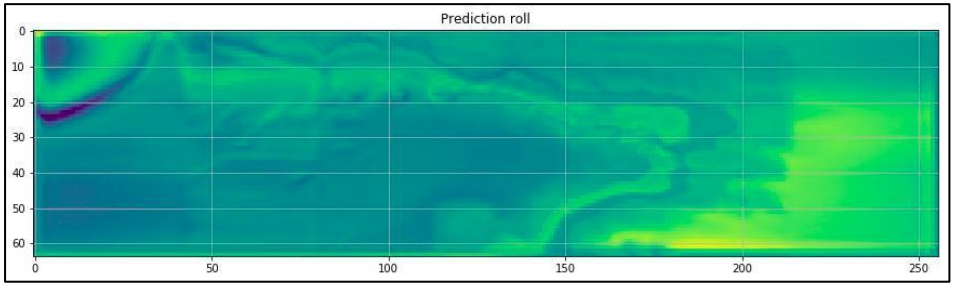
Prediction pitch_sin - after Epoch 9

Epochwise Output visualisation - roll

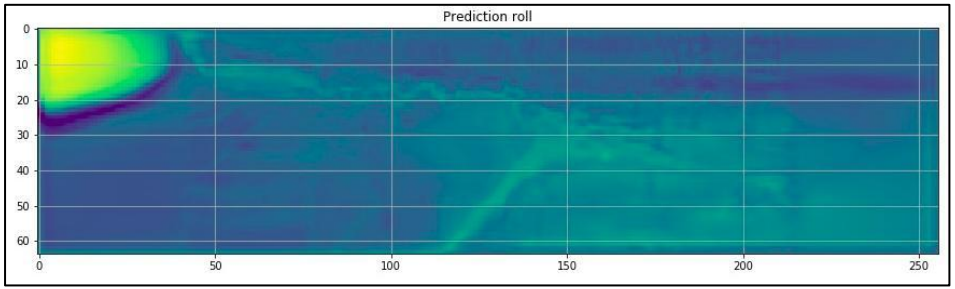
STAATLICH ANERKANNTE HOCHSCHULE



Ground Truth roll



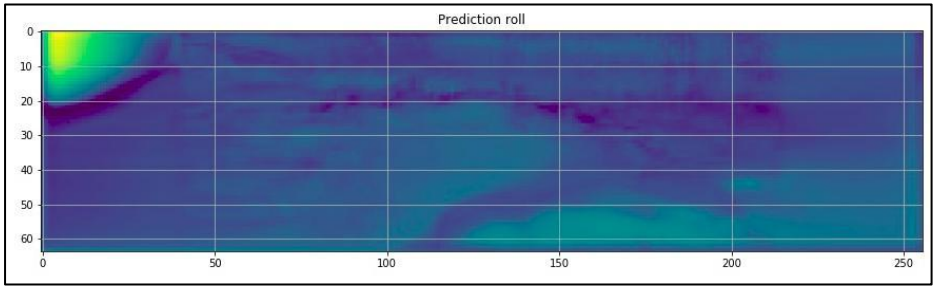
Prediction roll - after Epoch 0



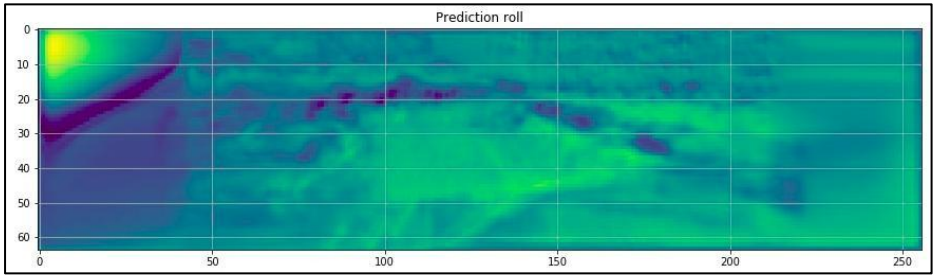
Prediction roll - after Epoch 1

Epochwise Output visualisation - roll

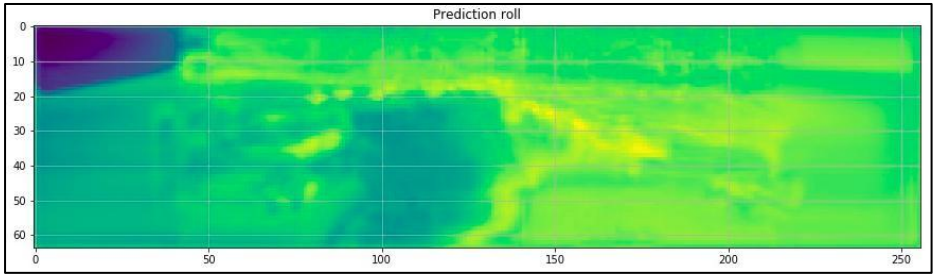
STAATLICH ANERKANNTE HOCHSCHULE



Prediction roll - after Epoch 2



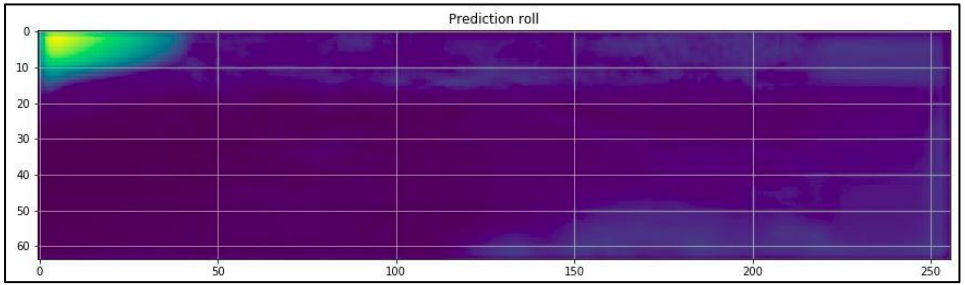
Prediction roll - after Epoch 3



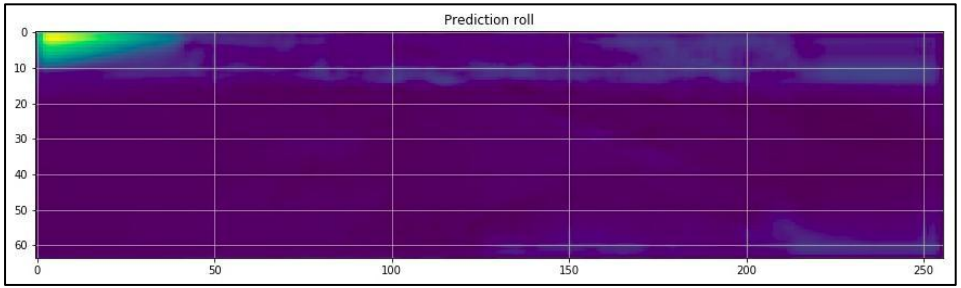
Prediction roll - after Epoch 4

Epochwise Output visualisation - roll

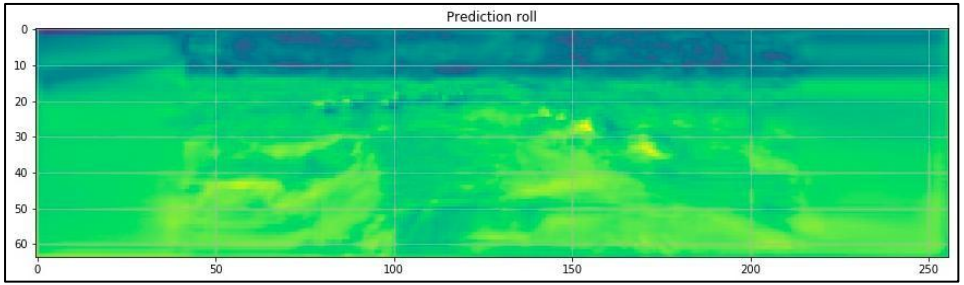
STAATLICH ANERKANNTE HOCHSCHULE



Prediction roll - after Epoch 5

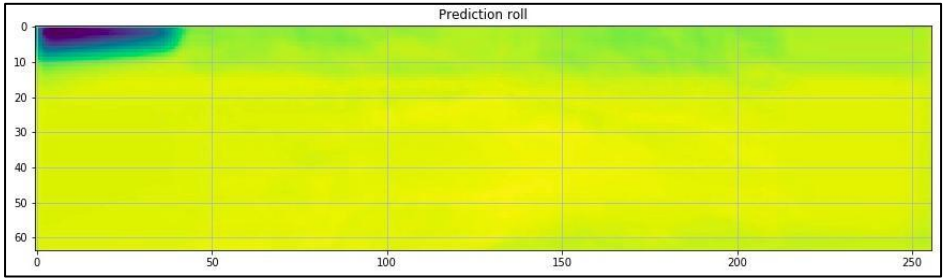


Prediction roll - after Epoch 6

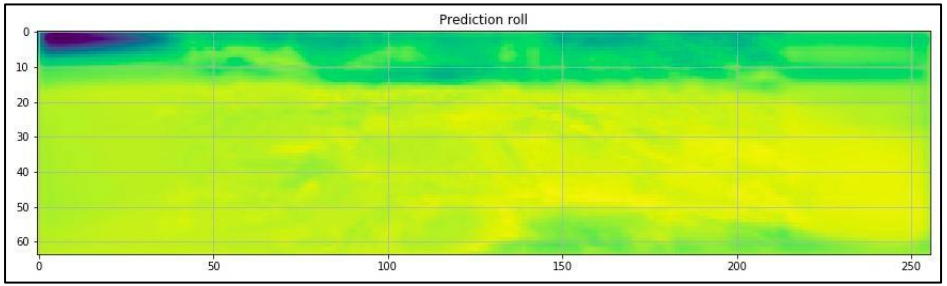


Prediction roll - after Epoch 7

Epochwise Output visualisation - roll



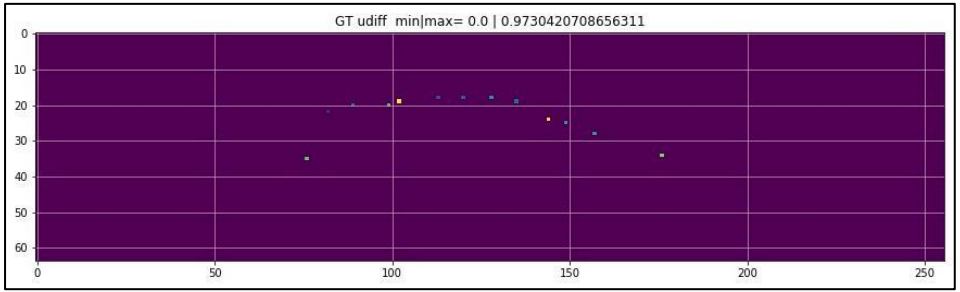
Prediction roll - after Epoch 8



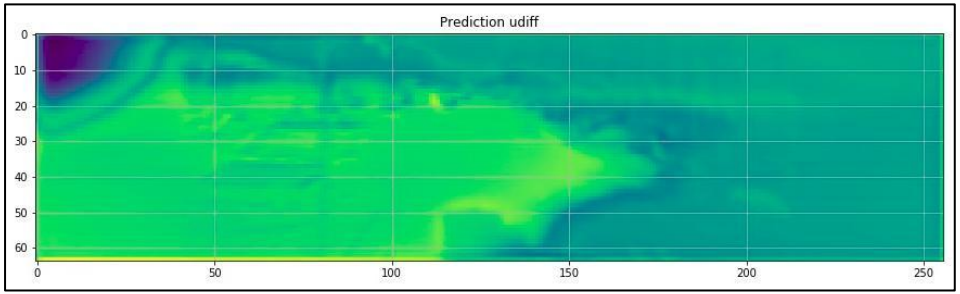
Prediction roll - after Epoch 9

Epochwise Output visualisation - udiff

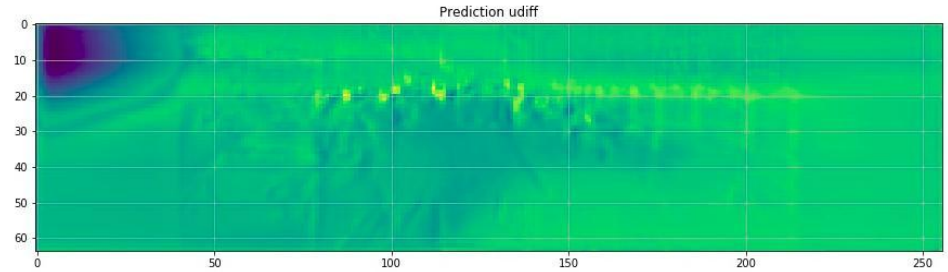
STAATLICH ANERKANNTE HOCHSCHULE



Ground Truth udiff



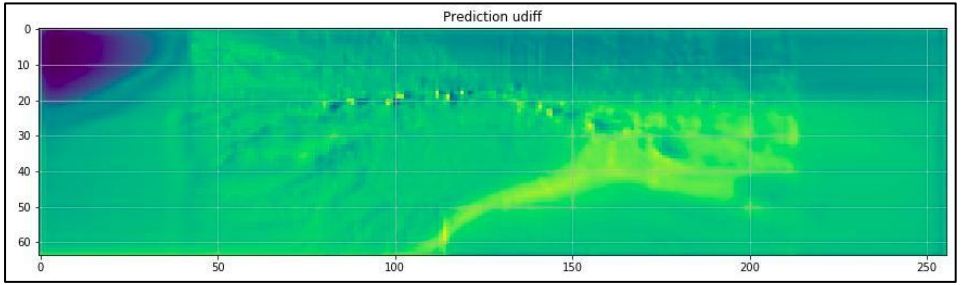
Prediction udiff - after Epoch 0



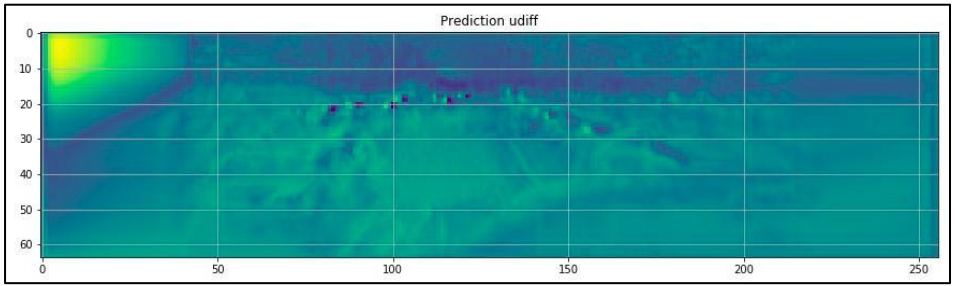
Prediction udiff - after Epoch 1

Epochwise Output visualisation - udiff

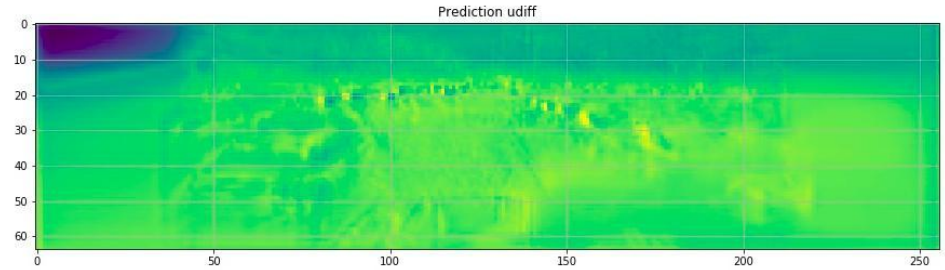
STAATLICH
ANERKANNTE
HOCHSCHULE



Prediction udiff - after Epoch 2



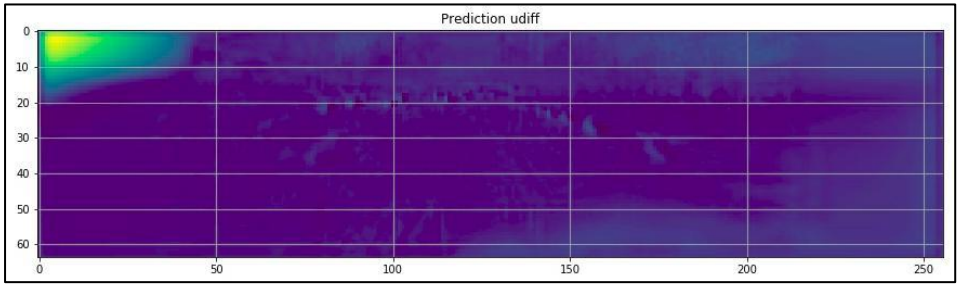
Prediction udiff - after Epoch 3



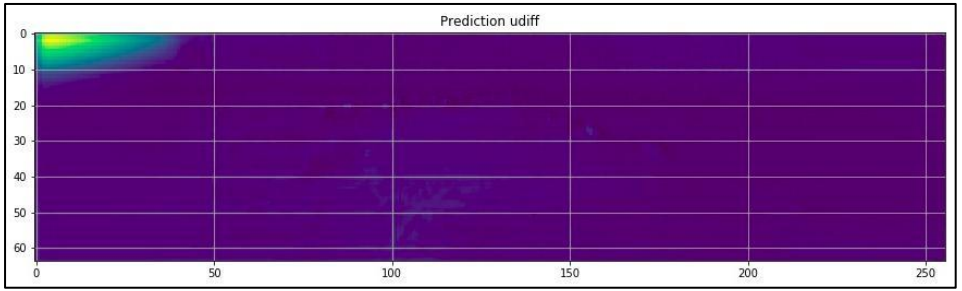
Prediction udiff - after Epoch 4

Epochwise Output visualisation - udiff

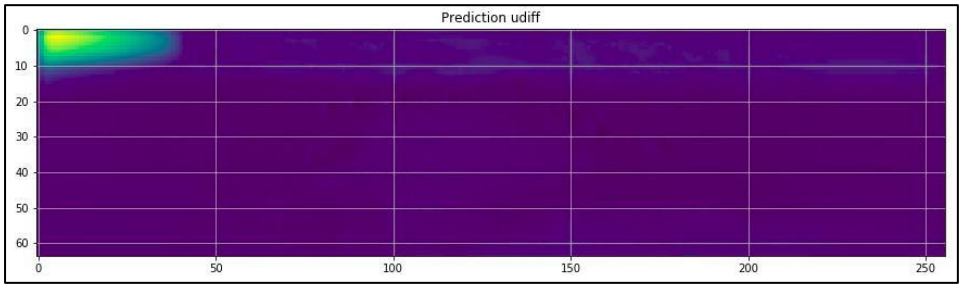
STAATLICH
ANERKANNTE
HOCHSCHULE



Prediction udiff - after Epoch 5



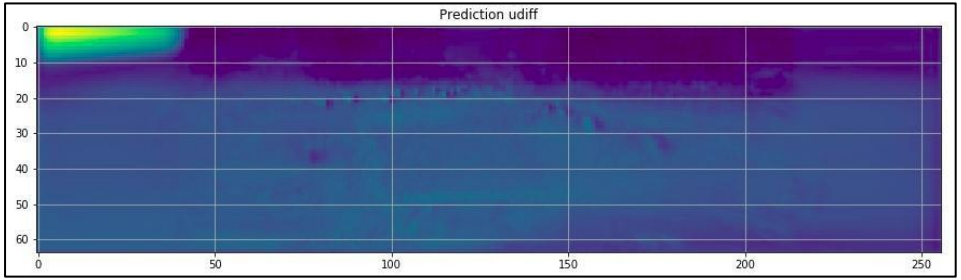
Prediction udiff - after Epoch 6



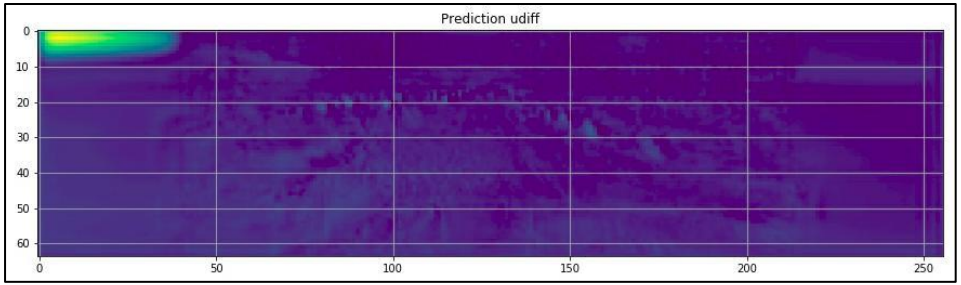
Prediction udiff - after Epoch 7

Epochwise Output visualisation - udiff

STAATLICH
ANERKANNTE
HOCHSCHULE



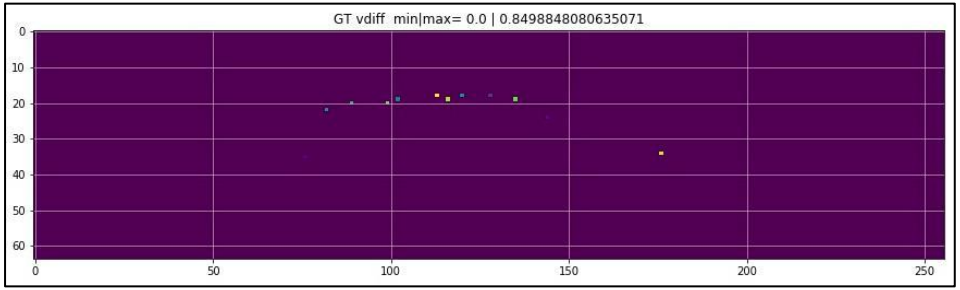
Prediction udiff - after Epoch 8



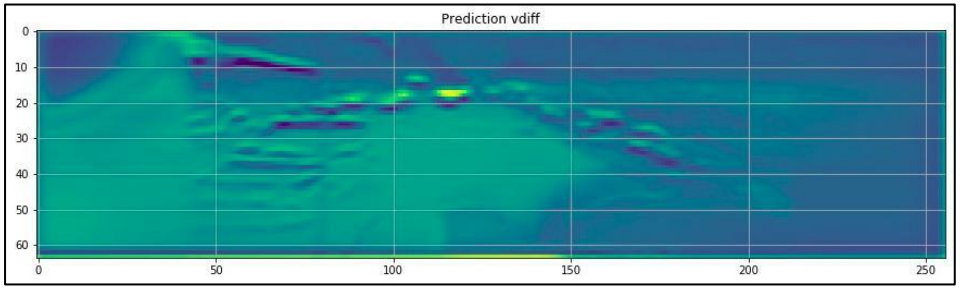
Prediction udiff - after Epoch 9

Epochwise Output visualisation - vdiff

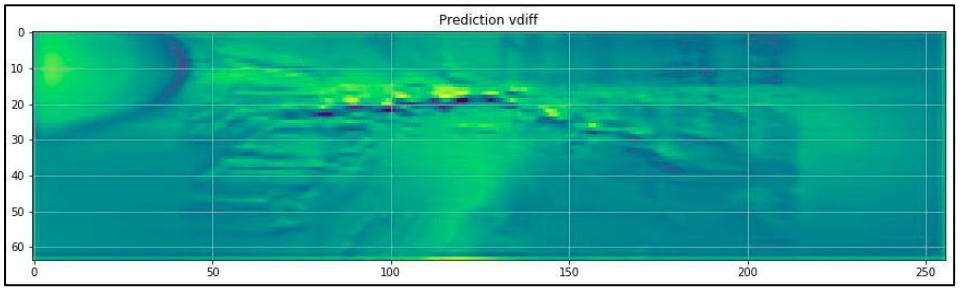
STAATLICH ANERKANNTE HOCHSCHULE



Ground Truth vdiff



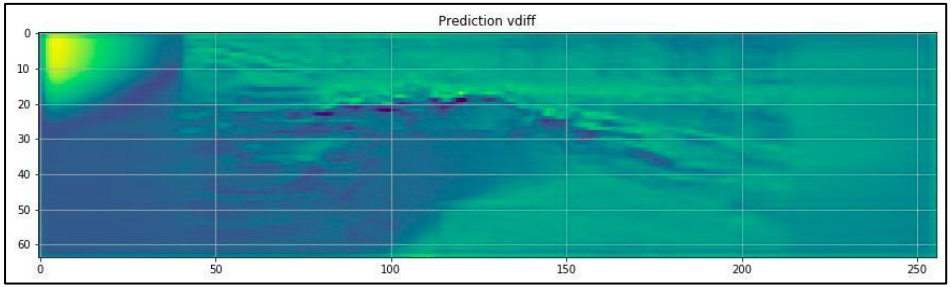
Prediction vdiff - after Epoch 0



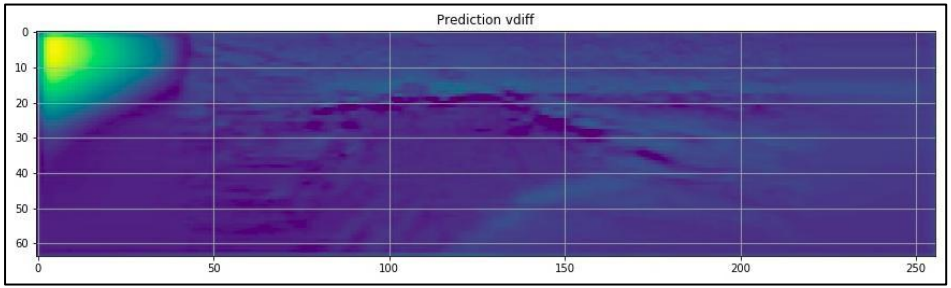
Prediction vdiff - after Epoch 1

Epochwise Output visualisation - vdiff

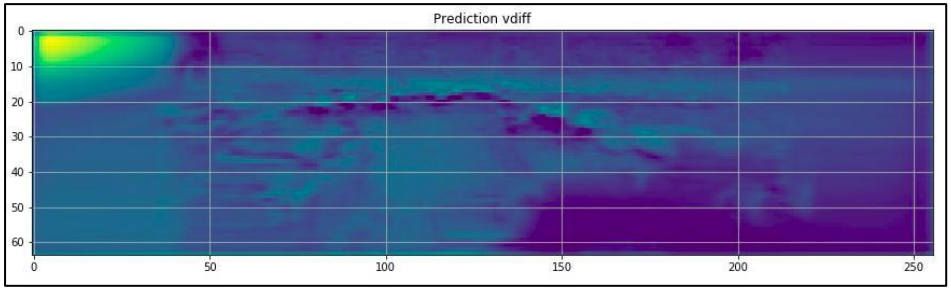
STAATLICH ANERKANNTE HOCHSCHULE



Prediction vdiff - after Epoch 2



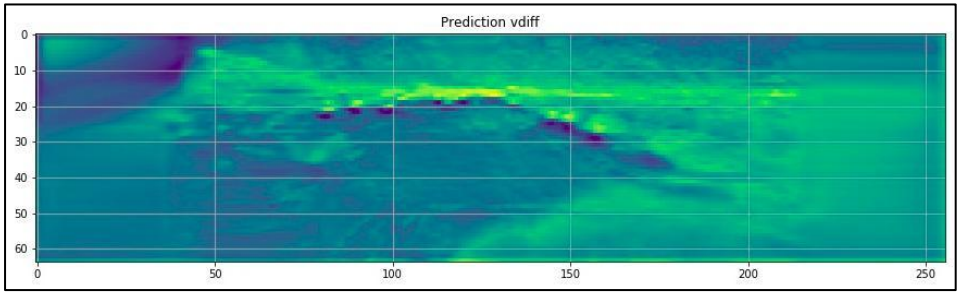
Prediction vdiff - after Epoch 3



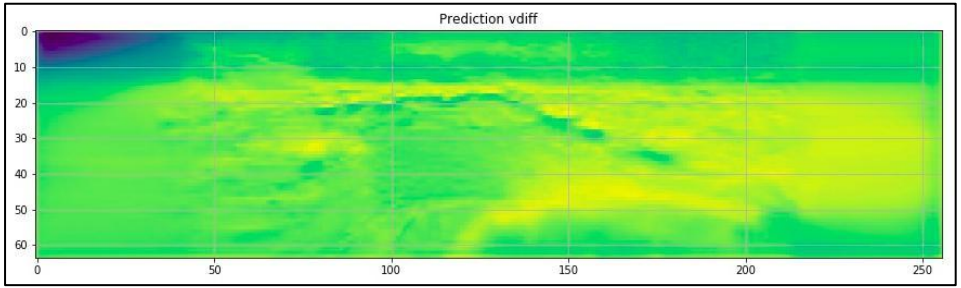
Prediction vdiff - after Epoch 4

Epochwise Output visualisation - vdiff

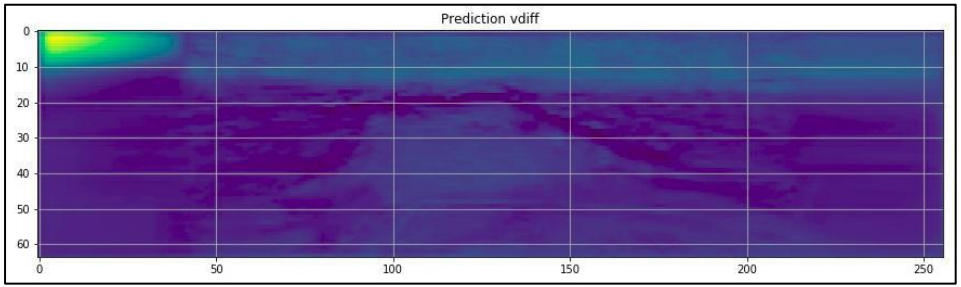
STAATLICH ANERKANNTE HOCHSCHULE



Prediction vdiff - after Epoch 5

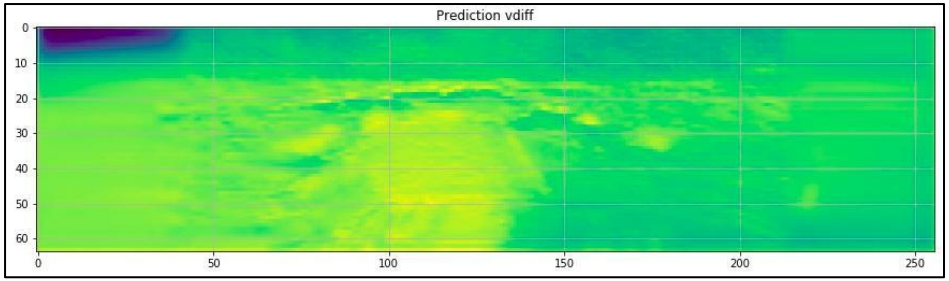


Prediction vdiff - after Epoch 6

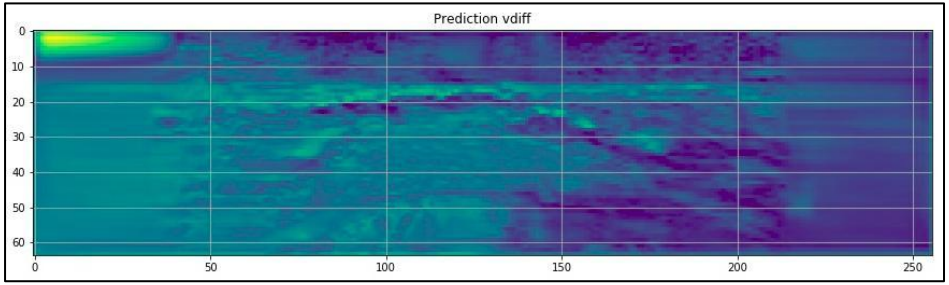


Prediction vdiff - after Epoch 7

Epochwise Output visualisation - vdiff



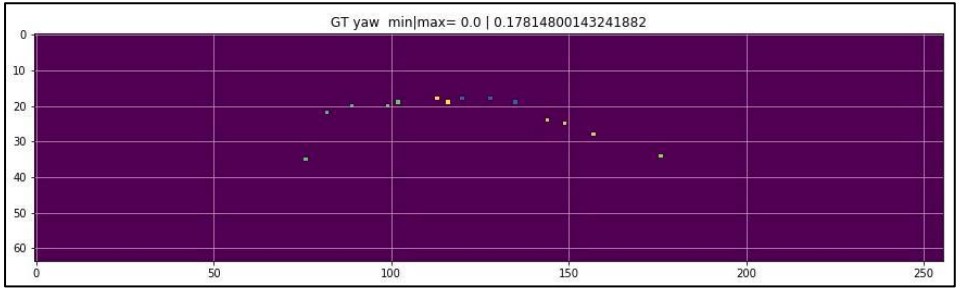
Prediction vdiff - after Epoch 8



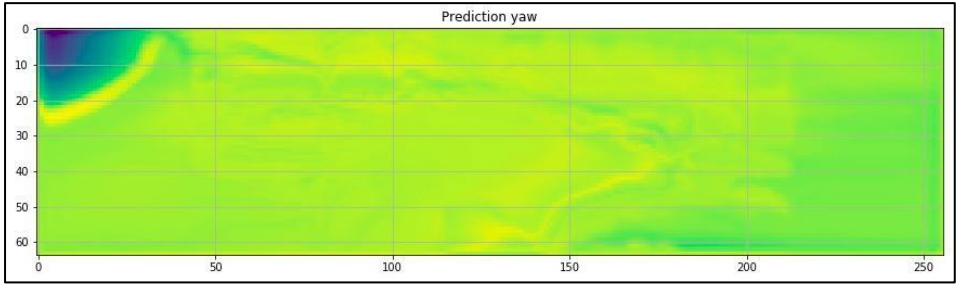
Prediction vdiff - after Epoch 9

Epochwise Output visualisation - yaw

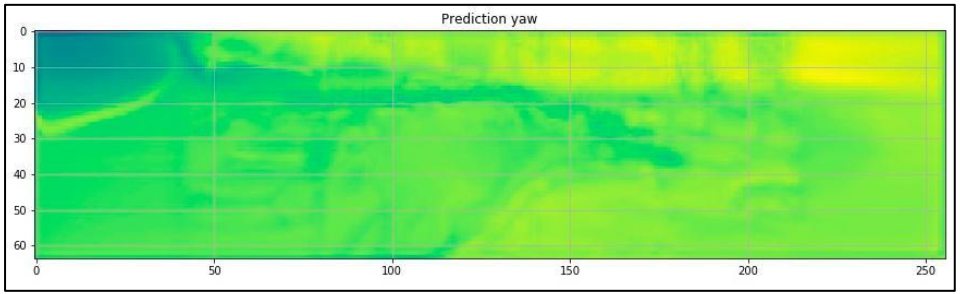
STAATLICH ANERKANNTE HOCHSCHULE



Ground Truth yaw



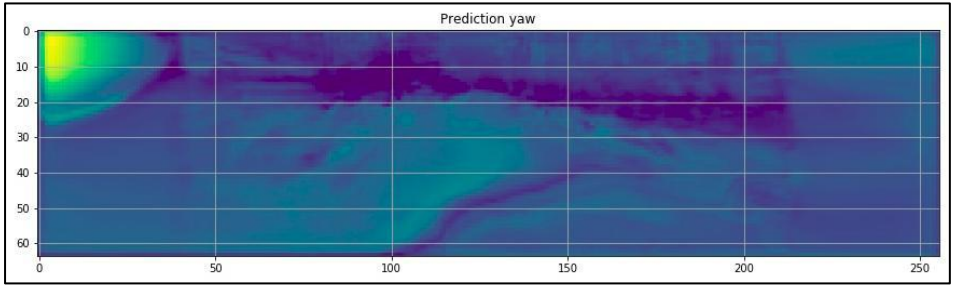
Prediction yaw - after Epoch 0



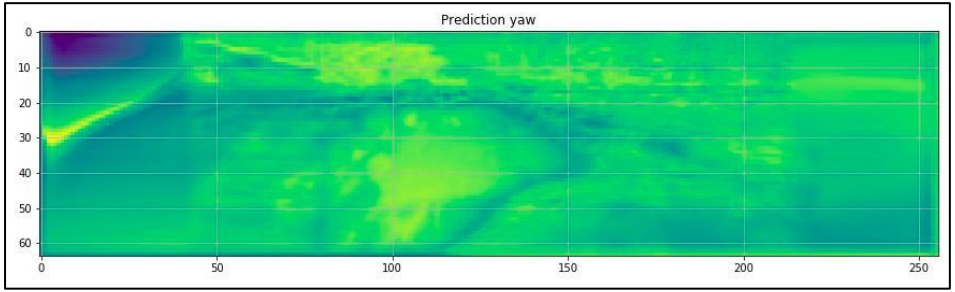
Prediction yaw - after Epoch 1

Epochwise Output visualisation - yaw

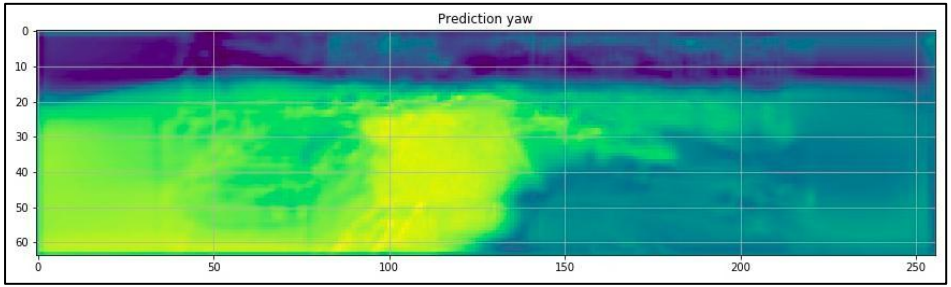
STAATLICH ANERKANNTE HOCHSCHULE



Prediction yaw - after Epoch 2



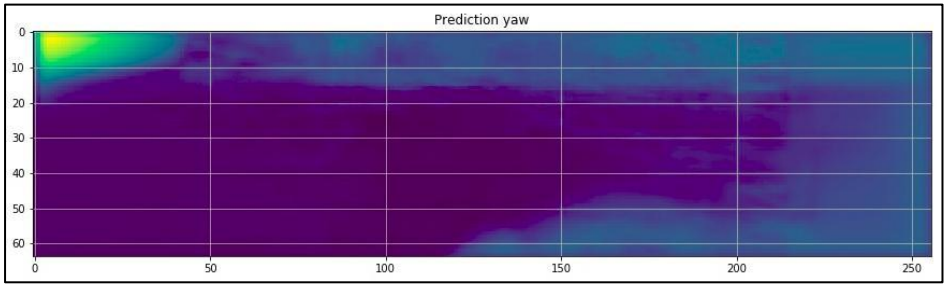
Prediction yaw - after Epoch 3



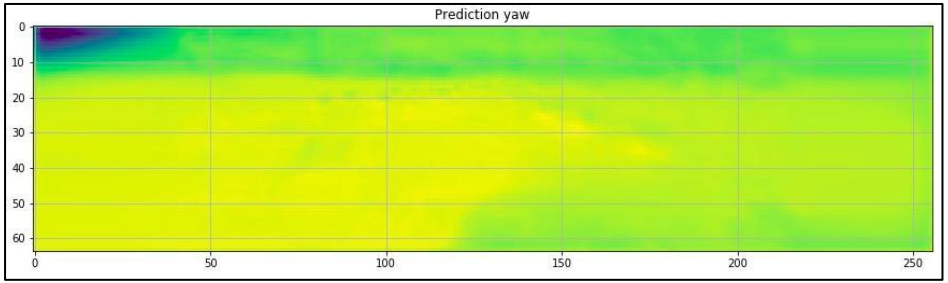
Prediction yaw - after Epoch 4

Epochwise Output visualisation - yaw

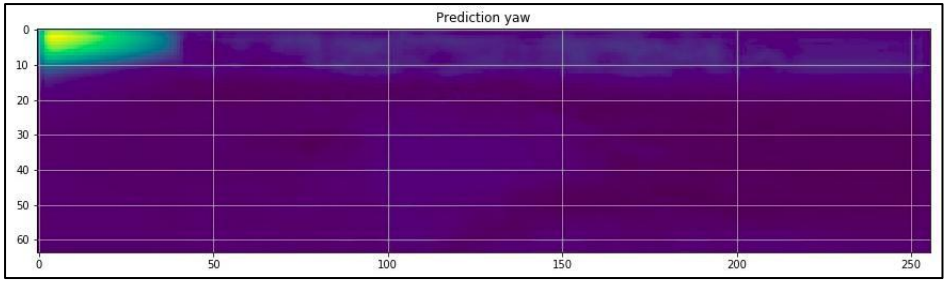
STAATLICH ANERKANNTE HOCHSCHULE



Prediction yaw - after Epoch 5



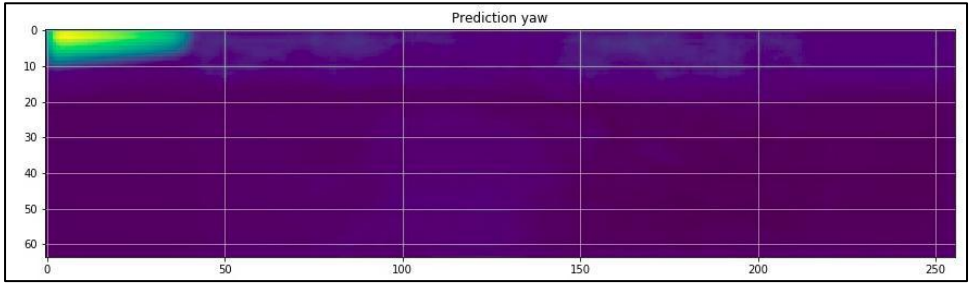
Prediction yaw - after Epoch 6



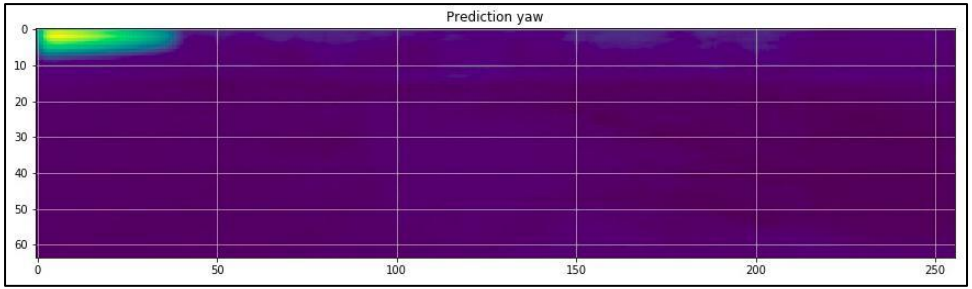
Prediction yaw - after Epoch 7

Epochwise Output visualisation - yaw

STAATLICH
ANERKANNTE
HOCHSCHULE



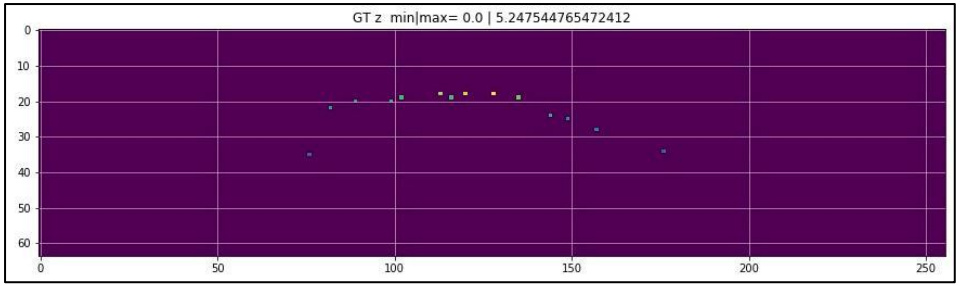
Prediction yaw - after Epoch 8



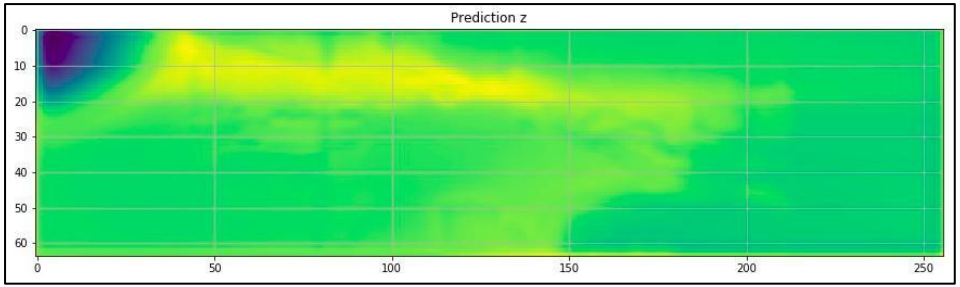
Prediction yaw - after Epoch 9

Epochwise Output visualisation - z

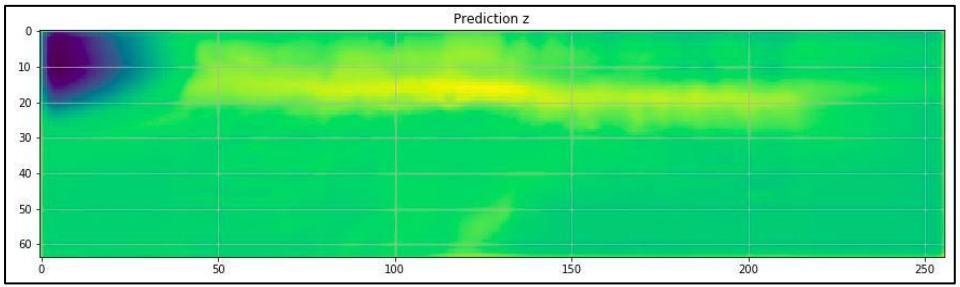
STAATLICH ANERKANNTE HOCHSCHULE



Ground Truth z



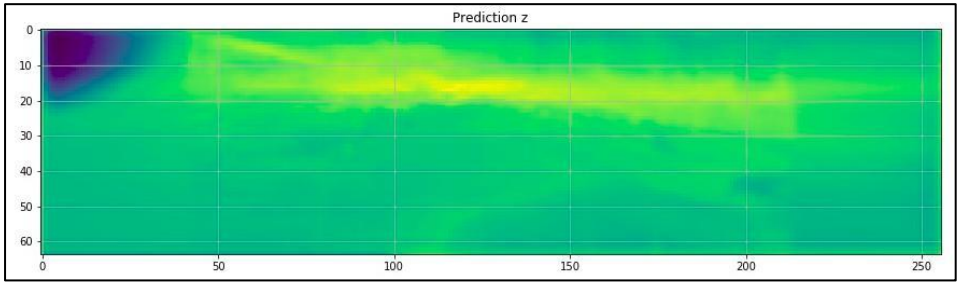
Prediction z - after Epoch 0



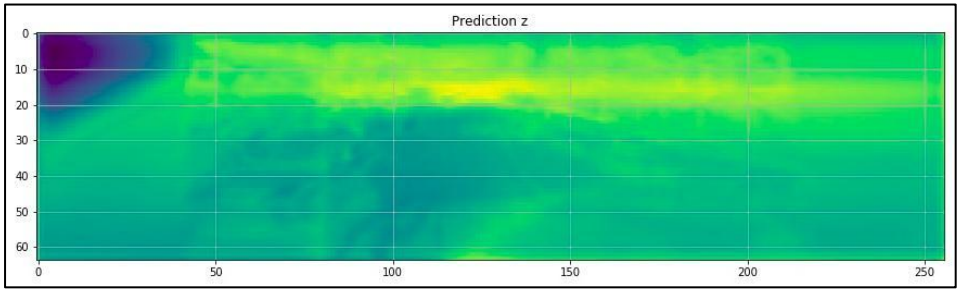
Prediction z - after Epoch 1

Epochwise Output visualisation - z

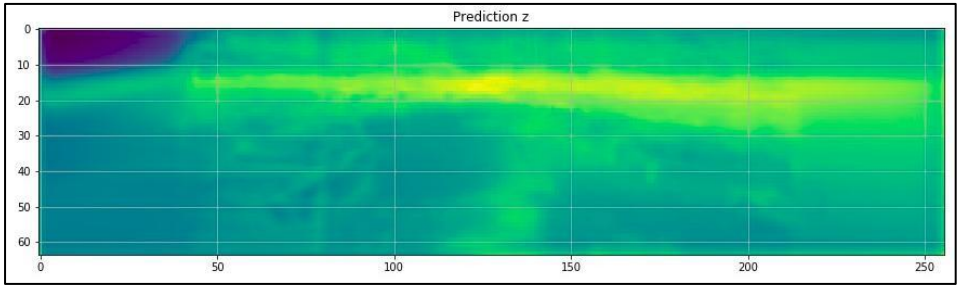
STAATLICH
ANERKANNTE
HOCHSCHULE



Prediction z - after Epoch 2



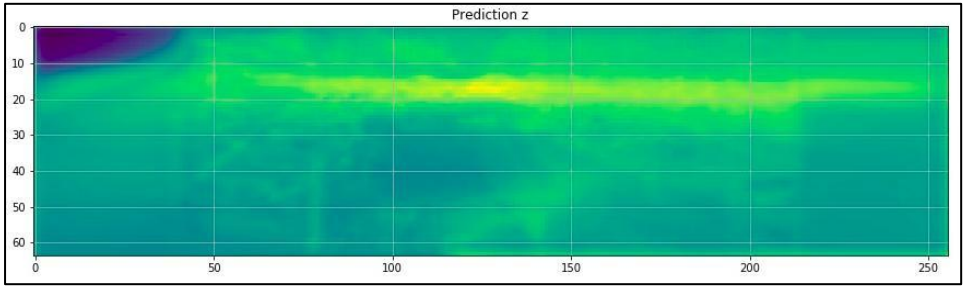
Prediction z - after Epoch 3



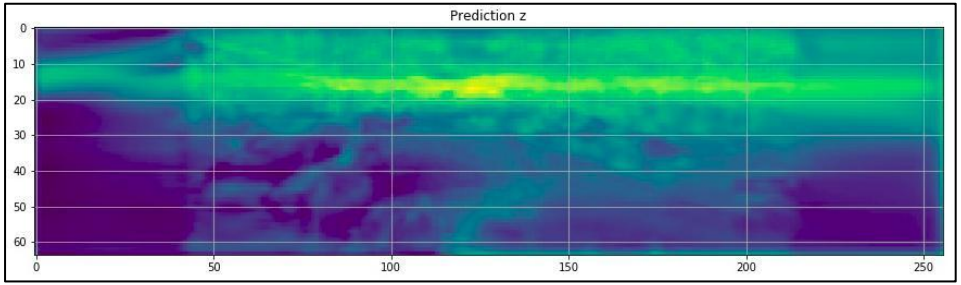
Prediction z - after Epoch 4

Epochwise Output visualisation - z

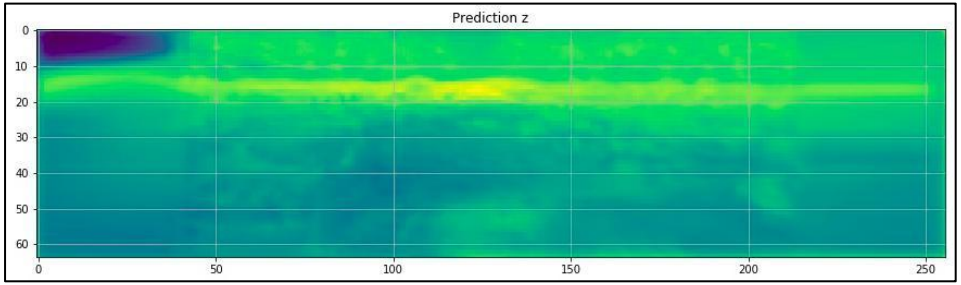
STAATLICH ANERKANNTE HOCHSCHULE



Prediction z - after Epoch 5



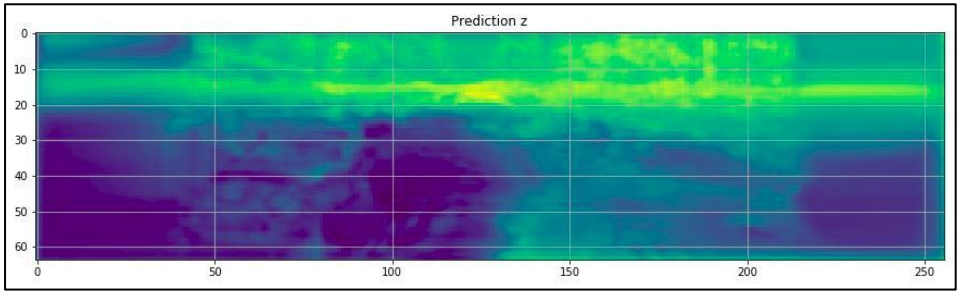
Prediction z - after Epoch 6



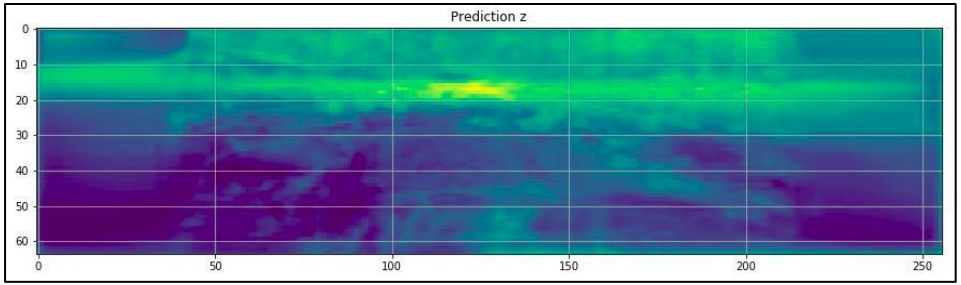
Prediction z - after Epoch 7

Epochwise Output visualisation - z

STAATLICH
ANERKANNTE
HOCHSCHULE



Prediction z - after Epoch 8



Prediction z - after Epoch 9

STAATLICH
ANERKANNTE
HOCHSCHULE

Thank You!

CenterNet High Level Block Diagram