

# Vision-based localization

Borna Bešić & Fabian Stiller

## Project goals

- Regress pose vector from AIS Perception Car image data using a deep neural network
- Compare two input models:
  - only front camera
  - six cameras covering 360°
- Smooth the trajectory given by the neural network



## Dataset

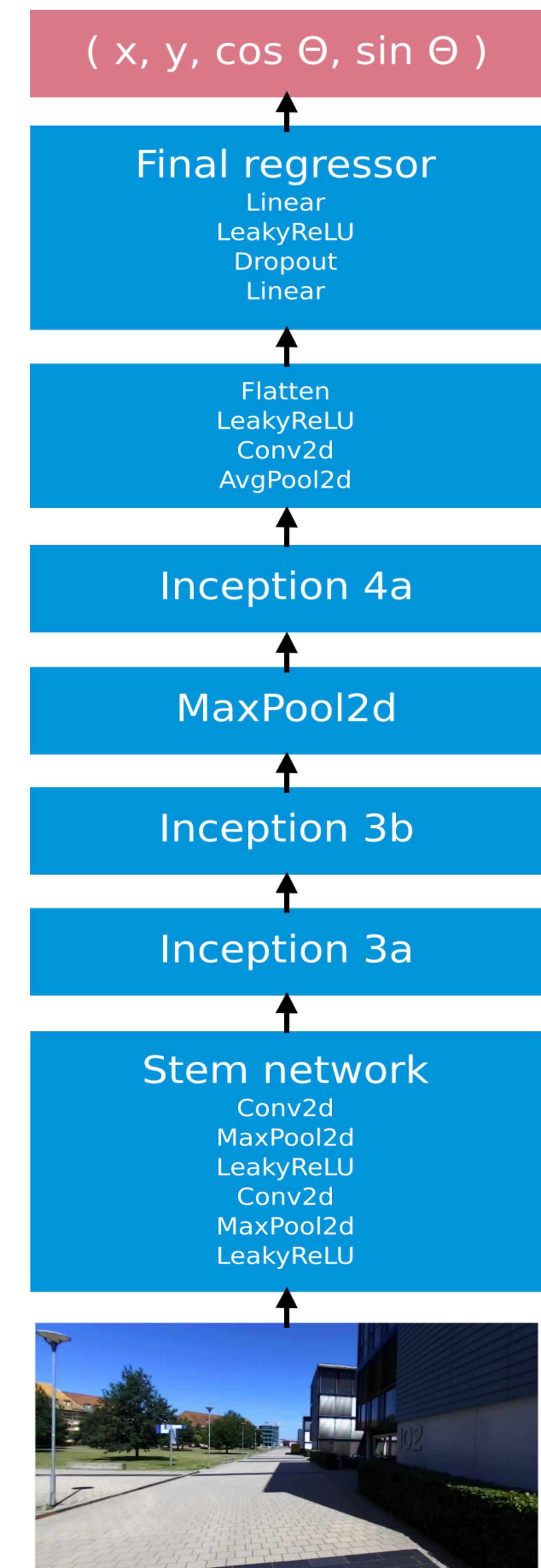
- ~ 9000 poses around Faculty of Engineering Campus with GPS data:
  - Universal Transverse Mercator (UTM) geographical coordinates
  - Quaternion orientation
- Spatial extent: 300 m x 300 m (90000 m<sup>2</sup>)
- For six cameras input model, images were concatenated along the channel axis resulting in a single image with 3 x 6 = 18 channels



## Neural network architecture

- PoseNet up to the first regressor
- Model size: ~ 9 MiB
- Loss function: combined position vector L2 loss and orientation vector L2 loss with a weighting factor  $\beta$

$$\begin{aligned} \mathbf{p} &= (\mathbf{x}, \mathbf{q}) \\ \mathbf{x} &= (x, y) \\ \mathbf{q} &= (\cos \theta, \sin \theta) \\ L(\mathbf{p}, \hat{\mathbf{p}}) &= \|\mathbf{x} - \hat{\mathbf{x}}\|_2 + \beta \left\| \mathbf{q} - \frac{\hat{\mathbf{q}}}{\|\hat{\mathbf{q}}\|_2} \right\|_2 \end{aligned}$$



## Experimental results

- Median position and orientation errors:
  - Front camera: 3.84 m, 3.97°
  - All cameras: 4.77 m, 2.71°

