Augmenting Computer Vision with expert knowledge and world-priors

dr. Nicola Strisciuglio, Assistant Professor, Computer Vision and Machine Learning

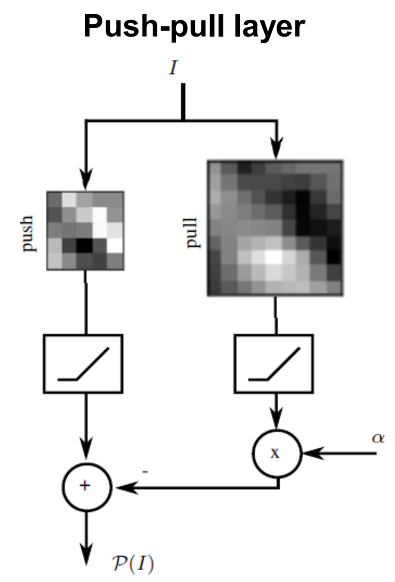
Vision and motivation: Currently, computer vision models tend to have a larger and lager parameter space. Training them requires (hundreds of) millions of images (often paired with text captions), and large, expensive and energy-hungry computation facilities. Large models are powerful as they are able to disentangle effective representations from a large number of images. However, they learn relying mainly on assumptions on data quality.

Our work is focused on including **explicit expert knowledge** or formalize **priors from real-world tasks** in the design and in the training process of convolutional architecture to facilitate the training of high-performance smaller models. We exploit neuroscientific findings and computational models, geometry-aware priors, and inspect the frequency characteristics of the data to exploit inductive bias and optimize training and application of CNN models.

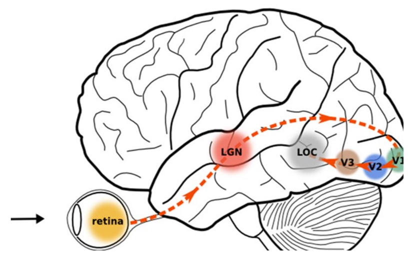
embedding expert/world knowledge

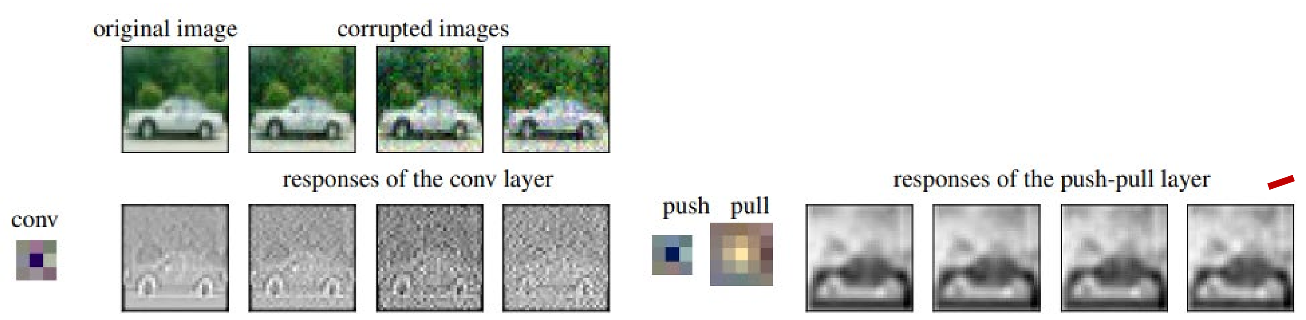
model robustness & generalization

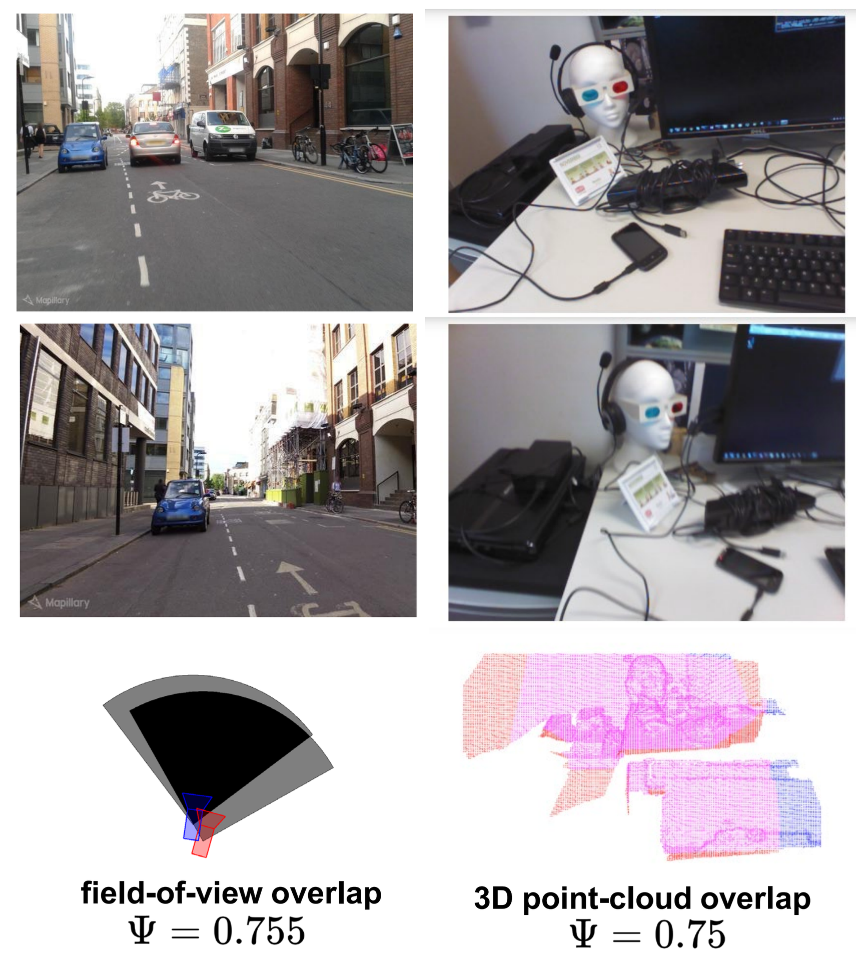
applications of CV



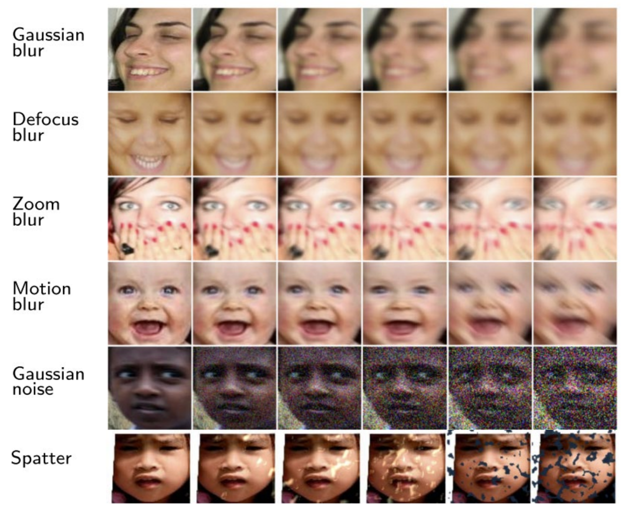
Brain-inspired components in CNNs. The visual system of the brain is able to carry out vision tasks efficiently and with great generalization abilities. We are inspired by expert knowledge of neuro-scientific findings to design and deploy computational models of neurons of the visual system into CNNs and computer vision models. We developed a novel Push-Pull layer, which implements a response inhibition mechanism in early stages of visual processing and enhances the robustness of CNNs to noisy and corrupted inputs. The Push-Pull layer combines a learnable convolutional kernel with another one of opposite polarity and larger support. The result is a network with the same number of parameters and improved generalization and robustness properties. We also study other mechanisms of neuronal response stabilization to regularize the training of CNNs.





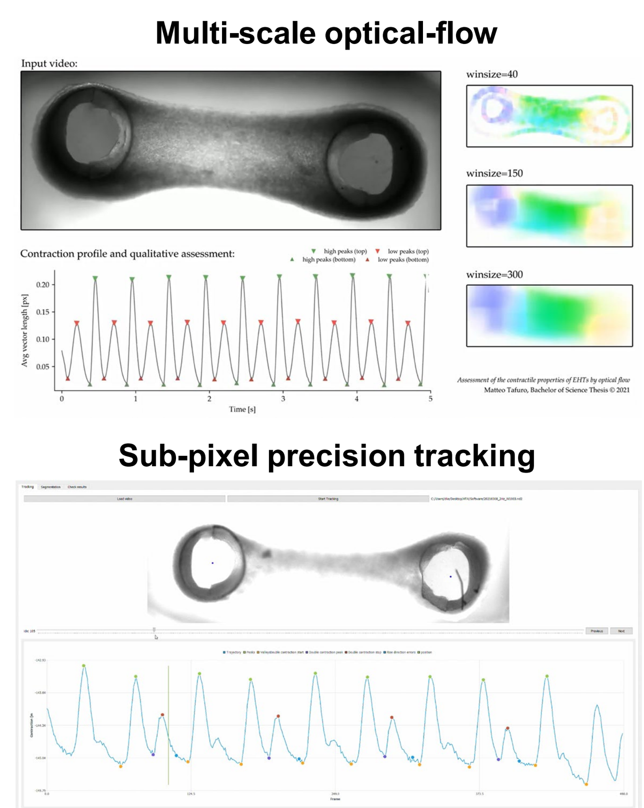


World- and camera-geometry aware priors for place recognition models. Traditional visual place recognition models are trained considering image similarity as a binary characteristic: two images are either depicting the same place or not. In practice, however, two images are *x%* similar. We exploit this prior concept and design a training strategy that involves world- and camera-geometry priors to estimate a graded similarity ground truth. A novel Generalized Contrastive Loss function allows to exploit the graded similarity information during the training stage, which only takes few hours, in contrast with existing approaches that need about 45 days on the same hardware. We achieve new SOTA results on the Mapillary Street Level Sequences data set.



Robustness to unknown perturbations. When deployed to work in real-world tasks, Computer Vision models are sensitive to perturbations of the input images that are not seen during training. Common perturbations include noise, various types of blur, illumination changes, elastic transformations, and so on, and considerably harm the performance of existing models. We study the factors that limit the generalization capabilities of existing models (e.g. shortcuts in the frequency domain) and investigate data- and architecture-based approaches to improve the robustness to unknown input perturbations. We aim at achieving models that can maintain high performance when deployed in scenarios with varying conditions.

Assessment of contraction properties of Engineered Heart Tissues. The BIOS and Applied Stem Cell lab at UT work on building artificial cardiac tissues from stem cells. The contraction properties of these tissues, cultured with different chemical compounds, are assessed, in order to eval their strength and applicability in large drug testing studies or to build artificial hearts. We support these studies by developing Computer Vision algorithms and dashboards for large-scale analysis of the contractile properties of these tissues. We deal with high-resolution in space and framerate, and tracking of very small movements by *sub-pixel precision* template matching and *multi-scale optical flow* algorithms. The algorithms are now used in the BIOS and Applied Stem Cell labs for assessment of tissue cultures.



Face emotion translation via motion-appearance disentanglement. We designed a GAN framework able to perform domain-translation of emotions in videos, maintaining face identity. We work on disentangling motion and appearance representations and ensure motion-consistency in the generated videos via an RNN-based generation of motion vectors.

Contacts
website: https://people.utwente.nl/n.strisciuglio
email: n.strisciuglio@utwente.nl

Thanks to all my students and collaborators for the brilliant work!

UNIVERSITY OF TWENTE.