Curse and bless of bias

Identifying, mitigating and inducing it in CV

Nicola Strisciuglio

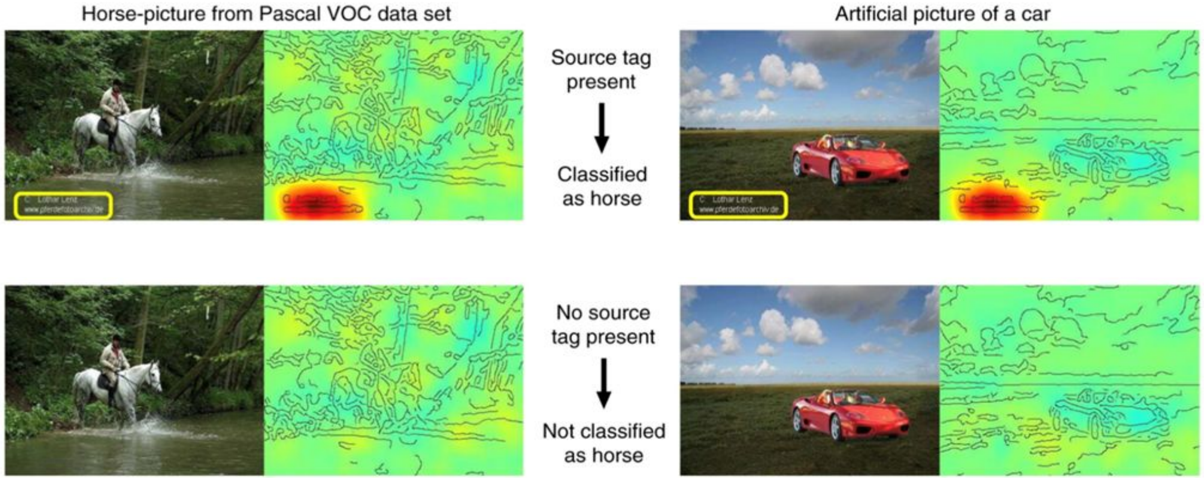
23 March 2024, ExploreCRS

UNIVERSITY OF TWENTE.

Contents

Identifying bias: Simplicity Bias and Shortcut Learning

Networks tend to **learn easy (less costly) solutions** to a problem [1]



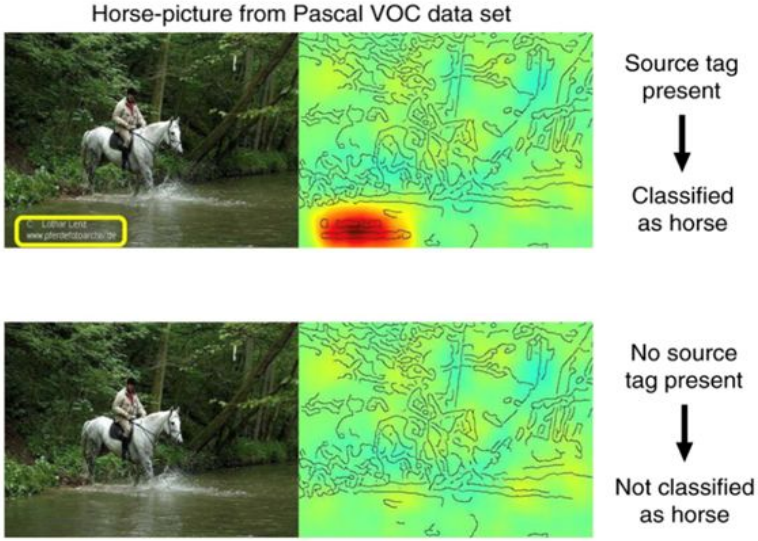
Lapuschkin, S. *et al.* **Unmasking Clever Hans predictors and assessing what machines really learn**. *Nat Commun* **10**, 1096 (2019).

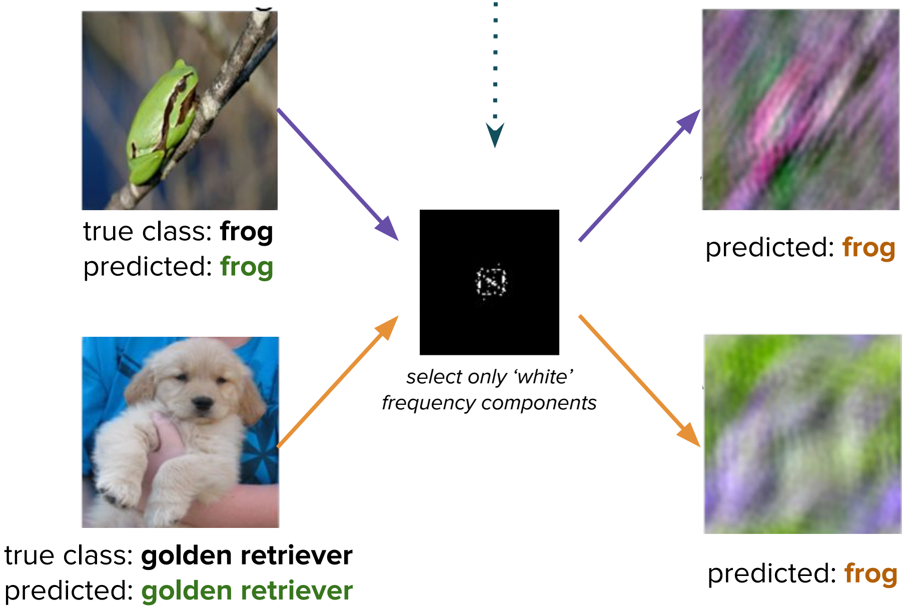
Shortcuts are decision rules based on ***spurious correlations between data and ground truth***, rather than on the correlation of semantic and task-related cues [2]

[1] Shah, H., *et al.* (2020). **The Pitfalls of Simplicity Bias in Neural Networks.** In NeurIPS 2020 (pp. 9573–9585).
[2] Geirhos, R., *et al.* (2020). **Shortcut learning in deep neural networks.** Nature Machine Intelligence, 2(11), 665–673.

Frequency (implicit) shortcuts

these can be computed - we designed an algorithm for ‘culling irrelevant frequencies’





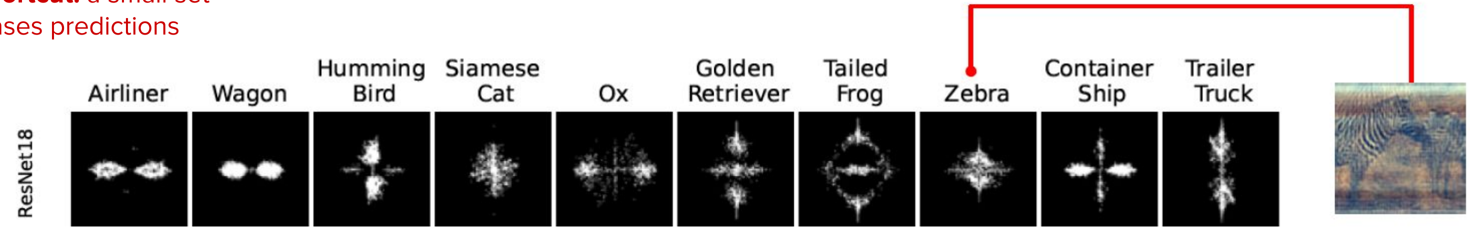
explicit, visible shortcuts

implicit, difficulty to ‘see’, frequency shortcuts

*S. Wang, R. Veldhuis, C. Brune, N. Strisciuglio, **What do neural networks learn in image classification? A frequency shortcut perspective**, ICCV 2023*

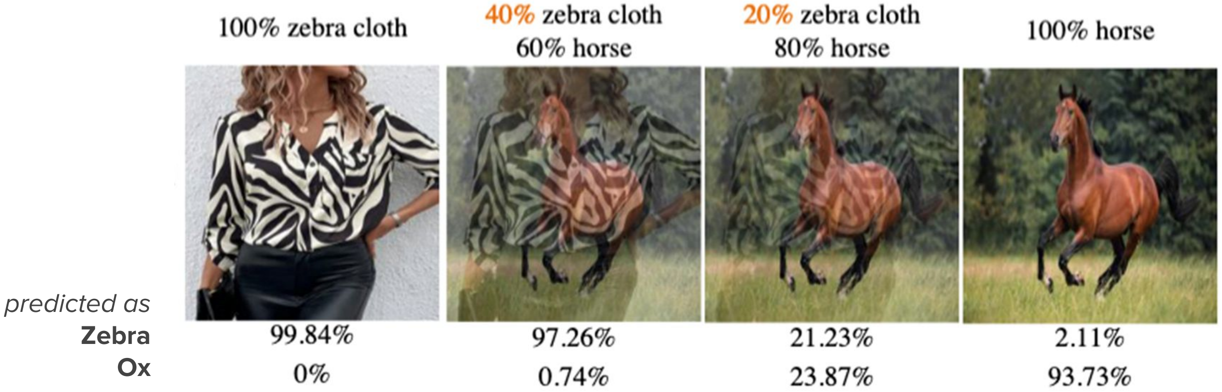
Frequency shortcuts and texture-bias of CNNs

affected by a frequency shortcut: a small set of frequency (texture) biases predictions



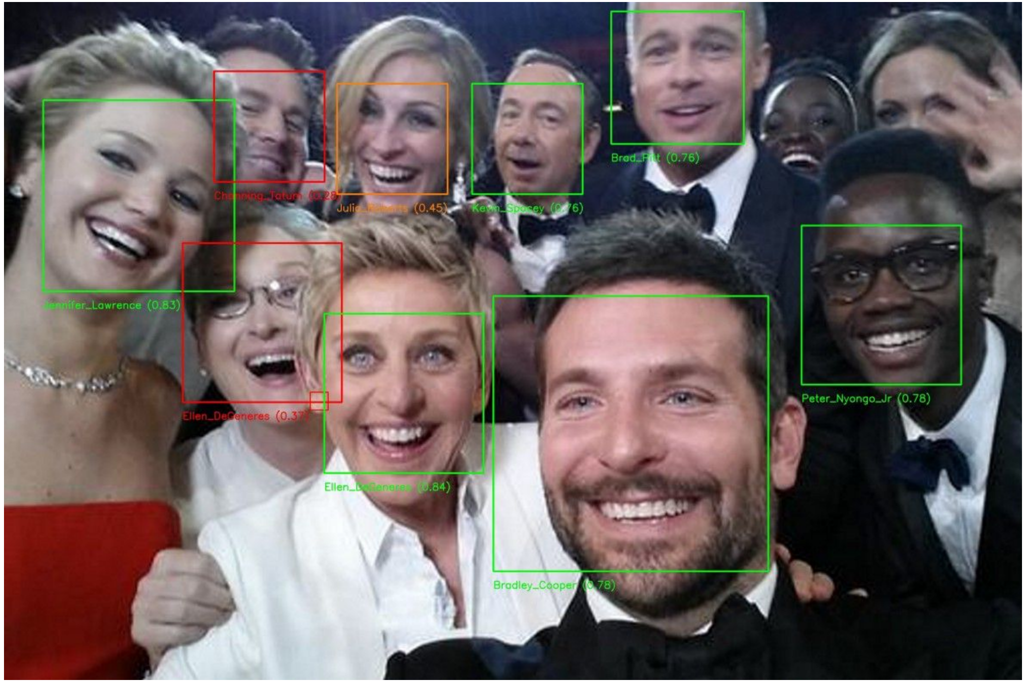
CNNs (trained on ImageNet) are texture-biased *(Geirhos et al, ICLR 2019)*: we can measure this bias

negative effects on generalization: not learning semantics

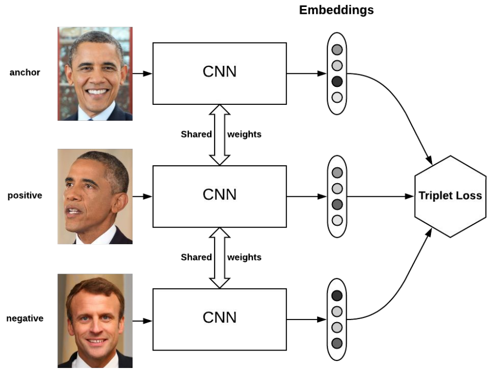


S. Wang, R. Veldhuis, C. Brune, N. Strisciuglio, ***What do neural networks learn in image classification? A frequency shortcut perspective***, ICCV 2023

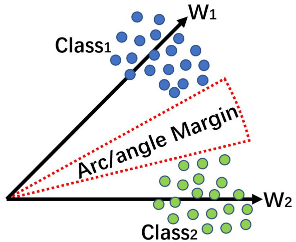
Mitigating bias in face analysis



FaceNet: minimize a triplet loss - push positive together and negative away



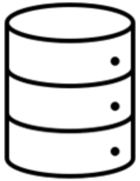
ArcLoss: faces of the same identity (classes) clustered together with high inter-class margin



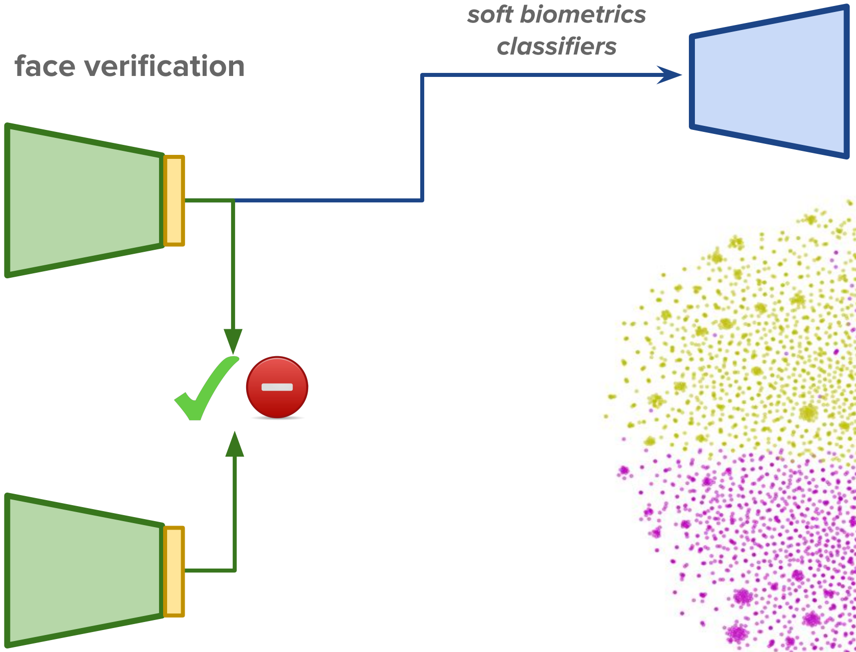
Problem!? Soft-biometrics are encoded in face recognition

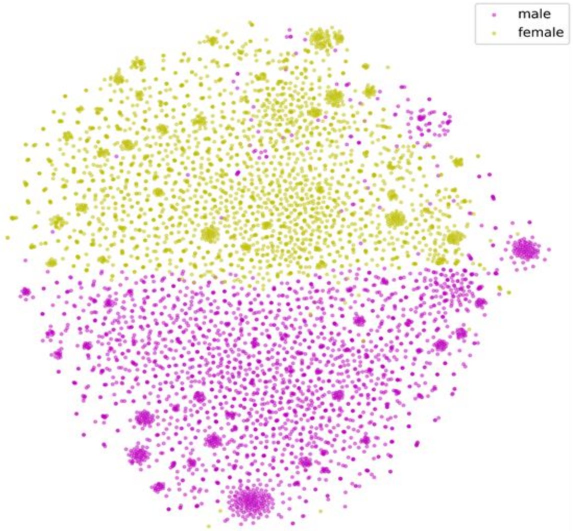
Un-bias the model from using soft biometrics => reduce (or trade-off) recognition by gender

Person to verify



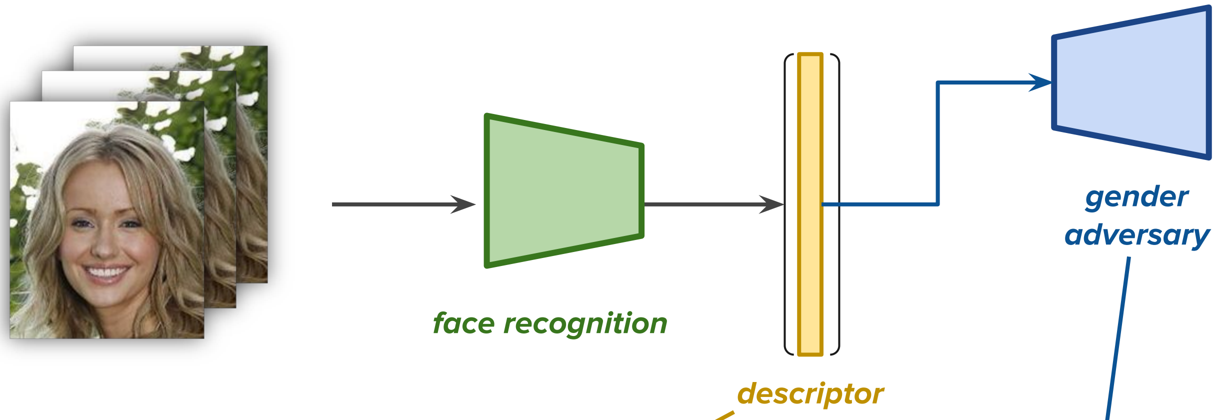
Identity database/
templates





Identity and gender distribution of
descriptors of LFW dataset samples

Reduce bias by a gender-privacy adversary



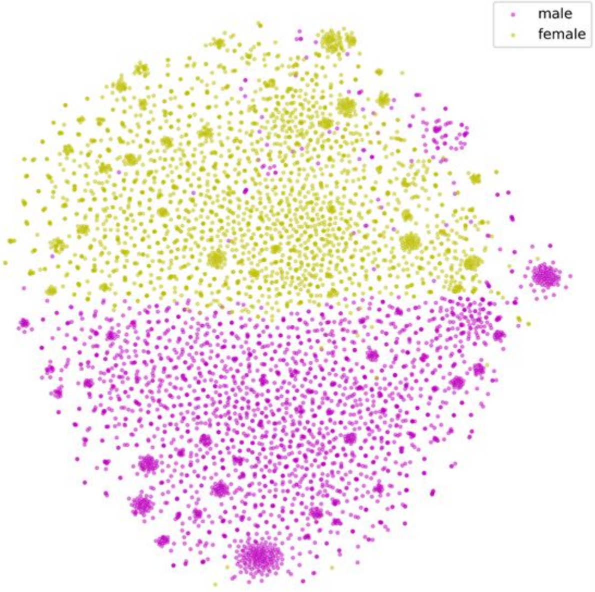
$$\mathcal{L} = AL(x) + (1-\alpha)\,|log(2) - CE(g(x))|$$

Arcloss
Identity

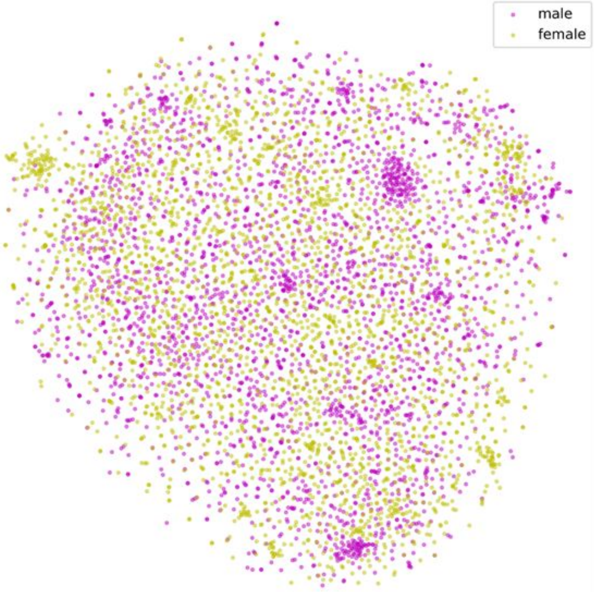
Cross Entropy
Gender-preserving

*Rezgui, Strisciuglio, Veldhuis, **under review***

Results: improved gender-privacy



arcface descriptors



arcface+gender adversary descriptors

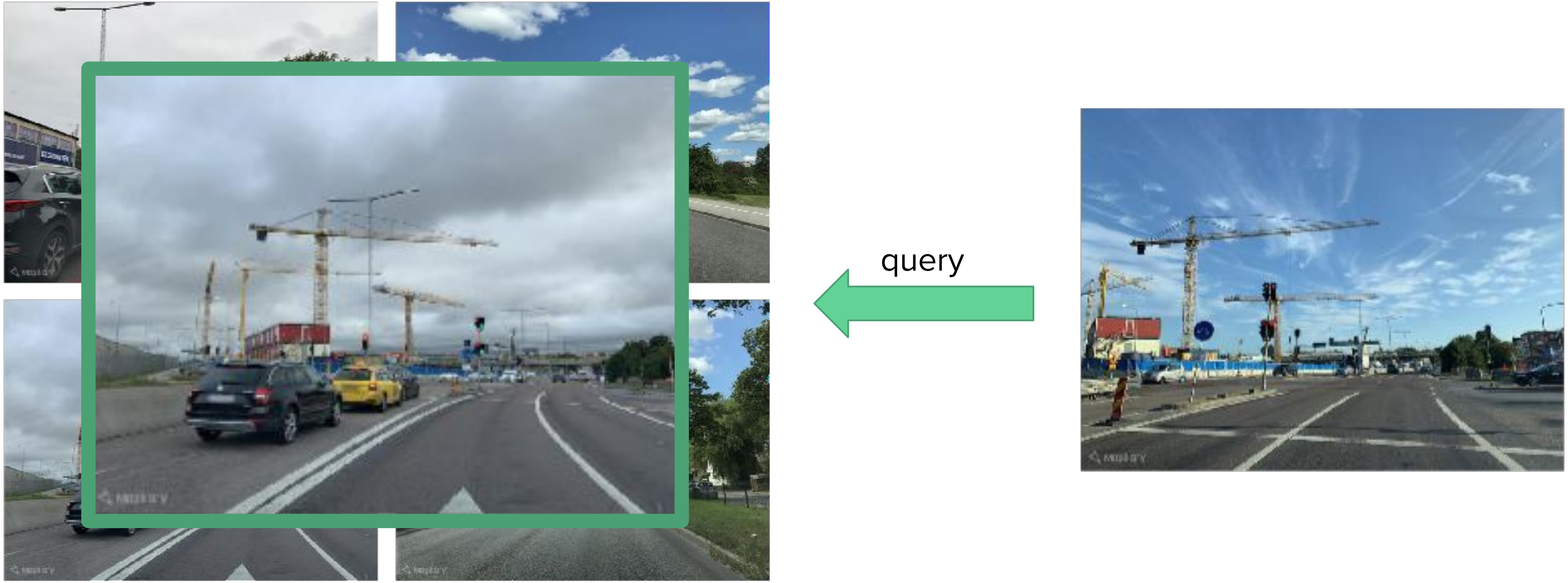
*Rezgui, Strisciuglio, Veldhuis, **under review***

Visual place recognition

Inductive bias via application-related priors

Objective:
recognize whether two images depict the same place, under seasonal and weather changes, time-of-the-day variations, etc.

Visual place recognition: **an Image Retrieval task**

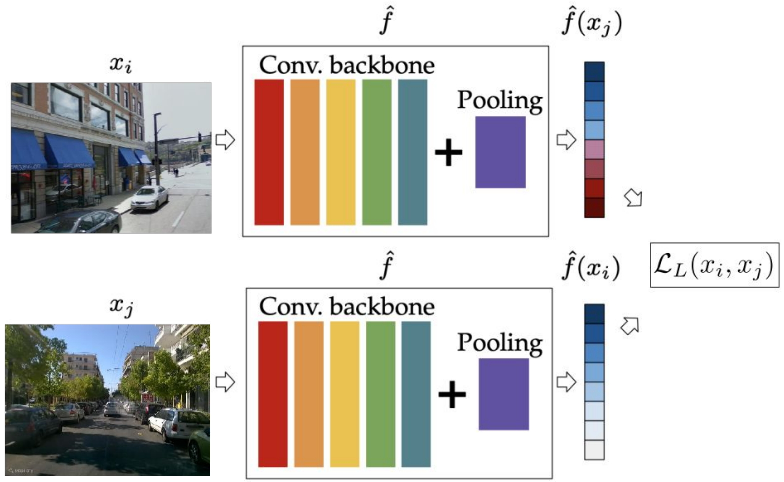


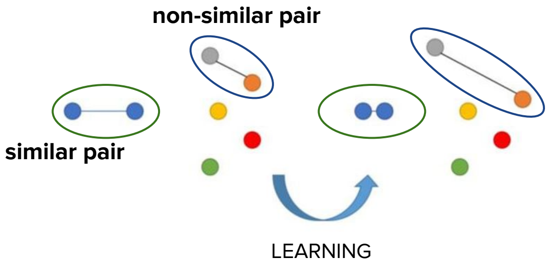
Contrastive learning of image descriptors

How
Train using image pairs (or triplets)

Objective

- Embed descriptor of **similar** images close together in a latent space
- ...descriptors of **non-similar** images are pushed away in a latent space





$$\mathcal{L}_{CL}(x_i, x_j) = \begin{cases} \frac{1}{2} d(\hat{f}(x_i), \hat{f}(x_j))^2, & \text{if } y_{i,j} = 1 \\ \frac{1}{2} \max(\tau - d(\hat{f}(x_i), \hat{f}(x_j)), 0)^2, & \text{if } y_{i,j} = 0 \end{cases}$$

Hadsell et al., ***Dimensionality Reduction by Learning an Invariant Mapping****, CVPR 2006*

Image similarity ground truth is *noisy*

***Rule in VPR**: images taken within 25 meters are of a similar place*

25.2m away
5.6m away

Image similarity as a *continuous* property

Similarity: $y_{i,j} \cancel{\in \{0, 1\}}$ needs to be reformulated $\longrightarrow \Psi_{i,j} \in [0, 1]$

~~binary similarity~~ continuous similarity

+++

5.6 m

+

24.7 m

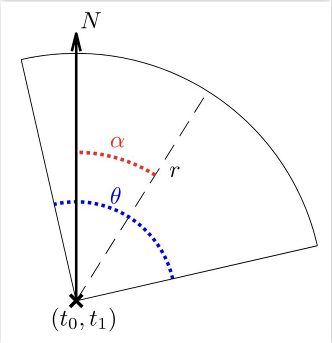
++

25.2 m

Leyva-Vallina, Strisciuglio, Petkov, **Data-efficient Large Scale Place Recognition with Graded Similarity Supervision,** CVPR 2023

Similarity via camera/scene geometry priors

Estimate the Field-of-View of the camera, using extra information in the data sets



r: radius
Θ: fov angle
α: direction

Mapillary Street Level Sequences (MSLS): GPS + compass angle

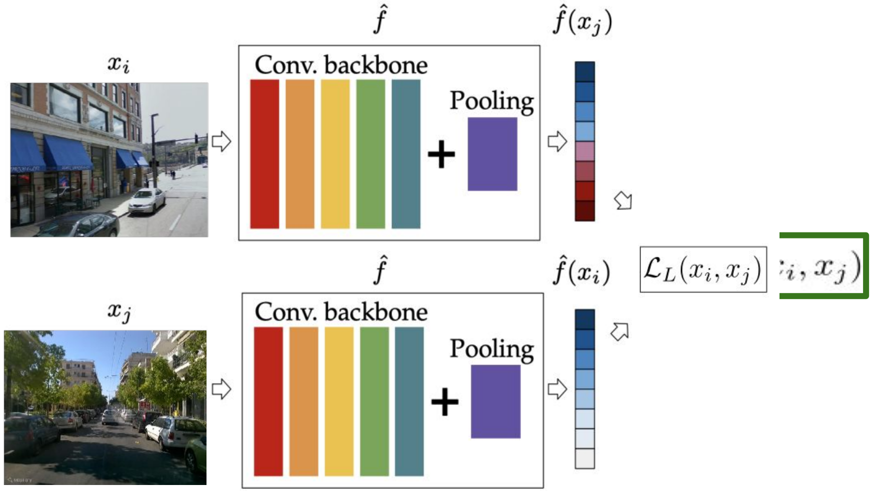
$\Psi = 0.755$

TrimBot2020 TB-Places: laser tracker + IMU (6dof camera pose)

$\Psi = 0.41$

Leyva-Vallina, Strisciuglio, Petkov, **Data-efficient Large Scale Place Recognition with Graded Similarity Supervision,** CVPR 2023

Generalized Contrastive Loss: induce continuous (pose) similarity prior into Contrastive Learning



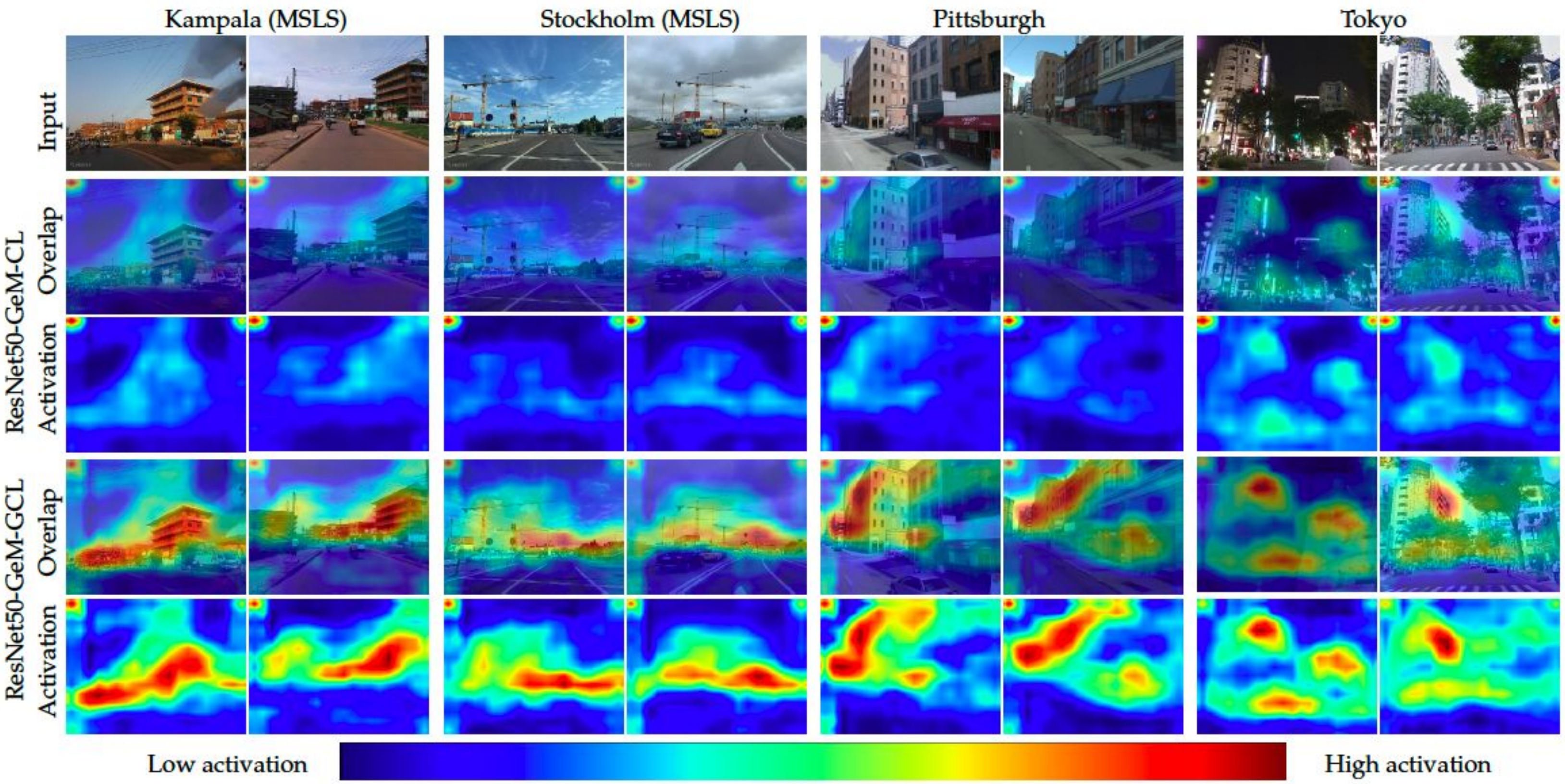
Generalized Contrastive Loss

$$\mathcal{L}_{GCL}(x_i, x_j) = \psi_{i,j} \frac{1}{2} d(f(x_i), f(x_j))^2 + (1 - \psi_{i,j}) \frac{1}{2} \max(\tau - d(f(x_i), f(x_j)), 0)^2$$

$$\mathcal{L}_{CL}(x_i, x_j) = \begin{cases} \frac{1}{2} d(\hat{f}(x_i), \hat{f}(x_j))^2 & \text{if } y_{i,j} = 1 \\ \frac{1}{2} \max(\tau - d(\hat{f}(x_i), \hat{f}(x_j)), 0)^2 & \text{if } y_{i,j} = 0 \end{cases}$$

continuous similarity

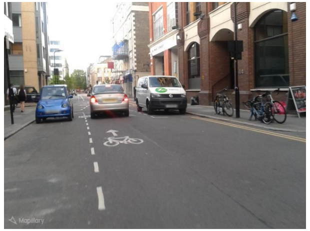
What the GCL looks at

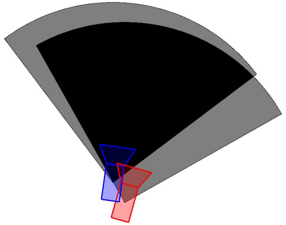


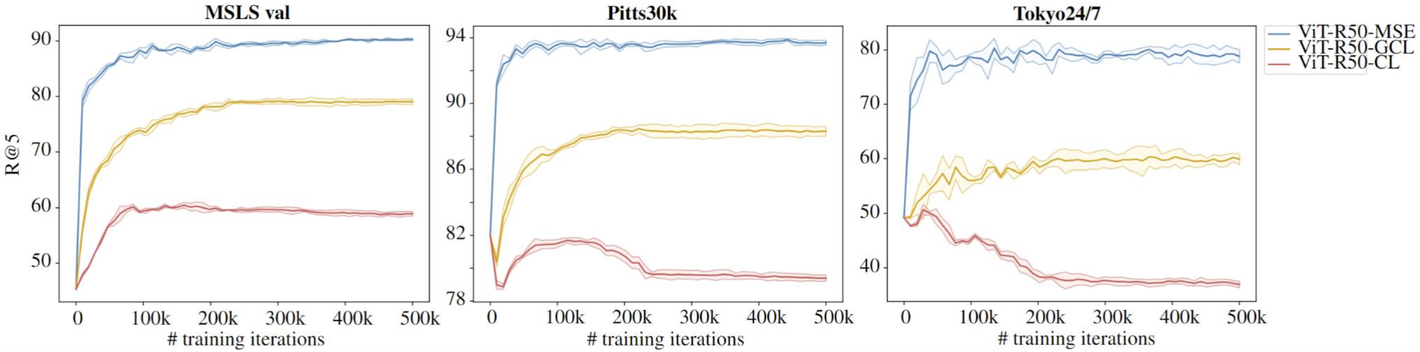
Leyva-Vallina, Strisciuglio, Petkov, **Data-efficient Large Scale Place Recognition with Graded Similarity Supervision,** CVPR 2023

What if? Learning VPR descriptors by regression

Descriptor distance as direct measure of fov overlap (image similarity)







data-efficiency and better performance

Leyva-Vallina, Strisciuglio, Petkov, **Regressing Transformers for Data-efficient Visual Place Recognition**, ICRA 2024

Summary and *take home* message

- Networks **may learn shortcuts** and/or biased descriptors (*bias in data*)
- **Prior knowledge** helps to robustify computer vision models
  - Visual system is robust to variations and generalizes well: use neurophysiology findings into CV model design (!)
  - Application-related priors (f.i. camera FoV overlap in VPR) also work
  - *... any ideas for priors ...*

Thank You!

Curse and bless of bias

Identifying, mitigating and inducing it in CV

Nicola Strisciuglio

n.strisciuglio@utwente.nl

UNIVERSITY OF TWENTE.