CVSSP Centre for Vision, Speech and Signal Processing

Deja-Vu: Weakly Supervised Feature Learning for Seasonal Invariance

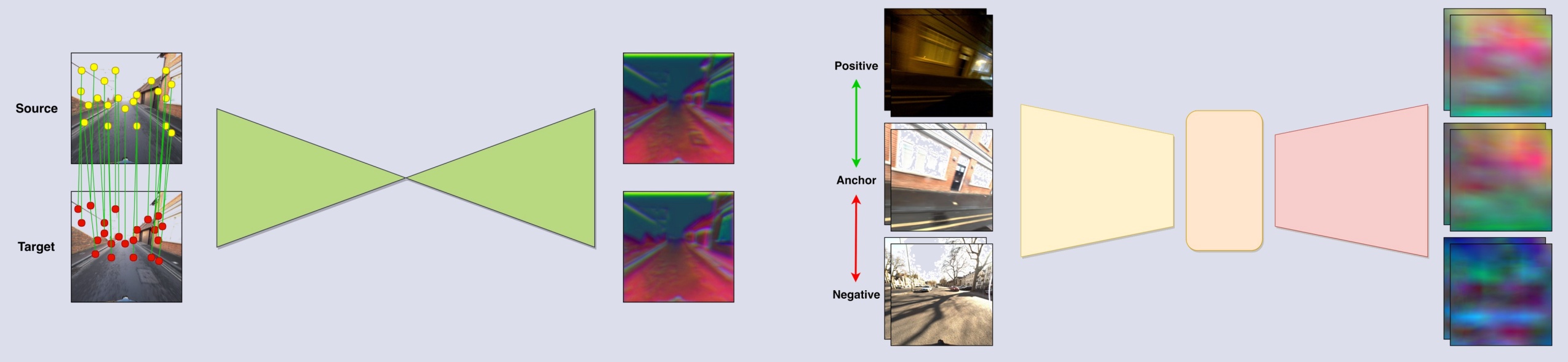
Jaime Spencer Martin (jaime.spencer@surrey.ac.uk) Richard Bowden (r.bowden@surrey.ac.uk) Simon Hadfield (s.hadfield.@surrey.ac.uk)

Centre for Vision, Speech and Signal Processing, Faculty of Engineering and Physical Sciences, University Of Surrey, Guildford, Surrey, GU2 7XH

UNIVERSITY OF SURREY

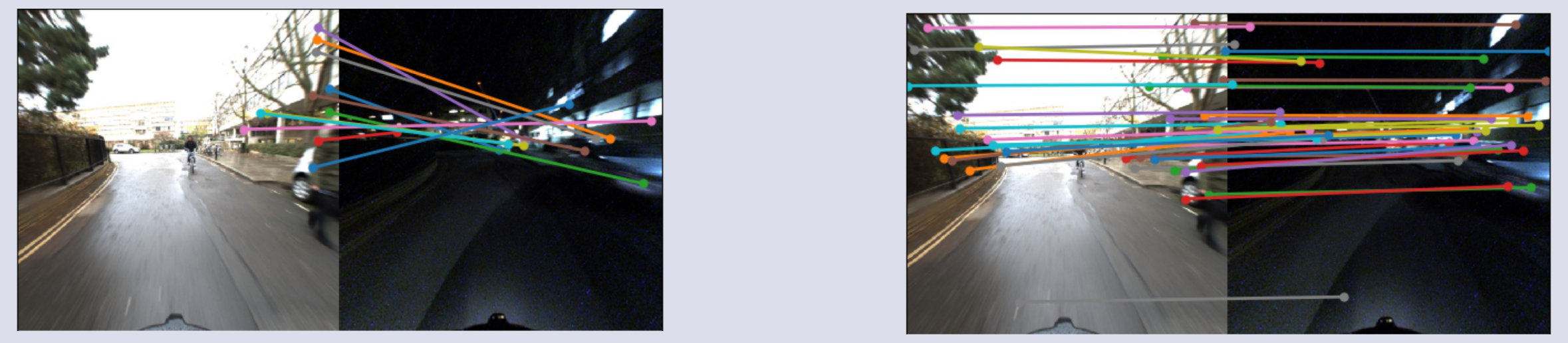
1 - Abstract

- We present **Deja-Vu**, a **weakly supervised** approach to **cross-seasonal** dense feature learning
- These features can be used in a **wide variety** of scenarios, serving as a **drop-in replacement** for existing solutions
- Contrary to most approaches, Deja-Vu instead requires **only rough alignment** indicating if a pair of images corresponds to the **same location or not**
- **Code** available at **github.com/jspenmar/DejaVu_Features**



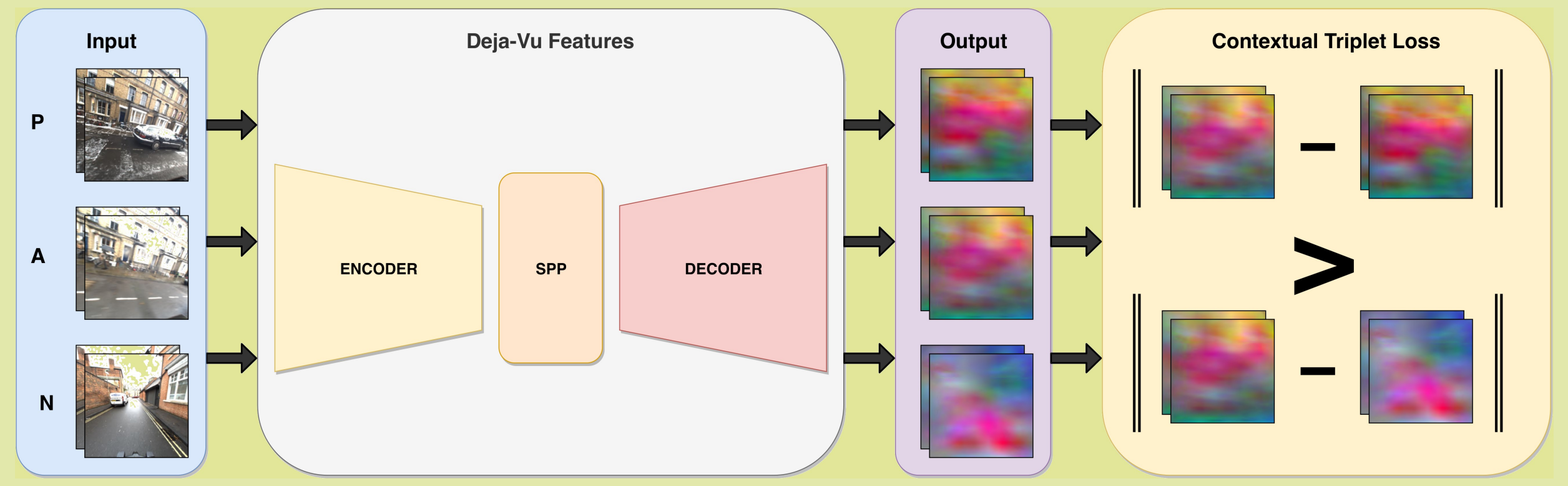
Traditional

Proposed



2 - Overview

- Encoder formed by **residual blocks + Spatial Pooling Pyramid** (32, 16, 8, 4)
- 3-block convolutional decoder with **skip connections**
- Maps $\boldsymbol{I} \in \mathbb{N}^{H \times W \times 3}$ to $\boldsymbol{F} \in \mathbb{R}^{H \times W \times n}$



3 - Links

Paper

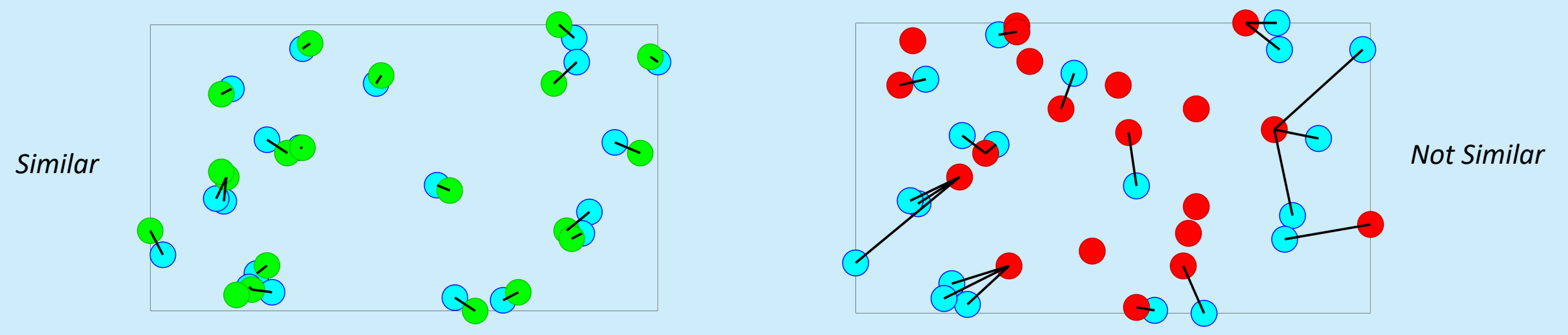
Code

LinkedIn

Other Research

4 - Contextual Similarity

- Determine **global "similarity"** between dense feature maps
- Each feature in I_1 should have **one matching feature** in I_2
- This match should be **significantly closer in embedding space** than all others



- Compute **distances** between all pair of features
- **Normalize** distances for each feature
- Compute **softmax** for each feature
- **Average** the highest softmax scores for all features

$$CX(\boldsymbol{F}_1, \boldsymbol{F}_2) = \frac{1}{N} \sum_{\boldsymbol{p}_1} \max \widetilde{S}(\boldsymbol{p}_1)$$

5 - Contextual Triplet Loss

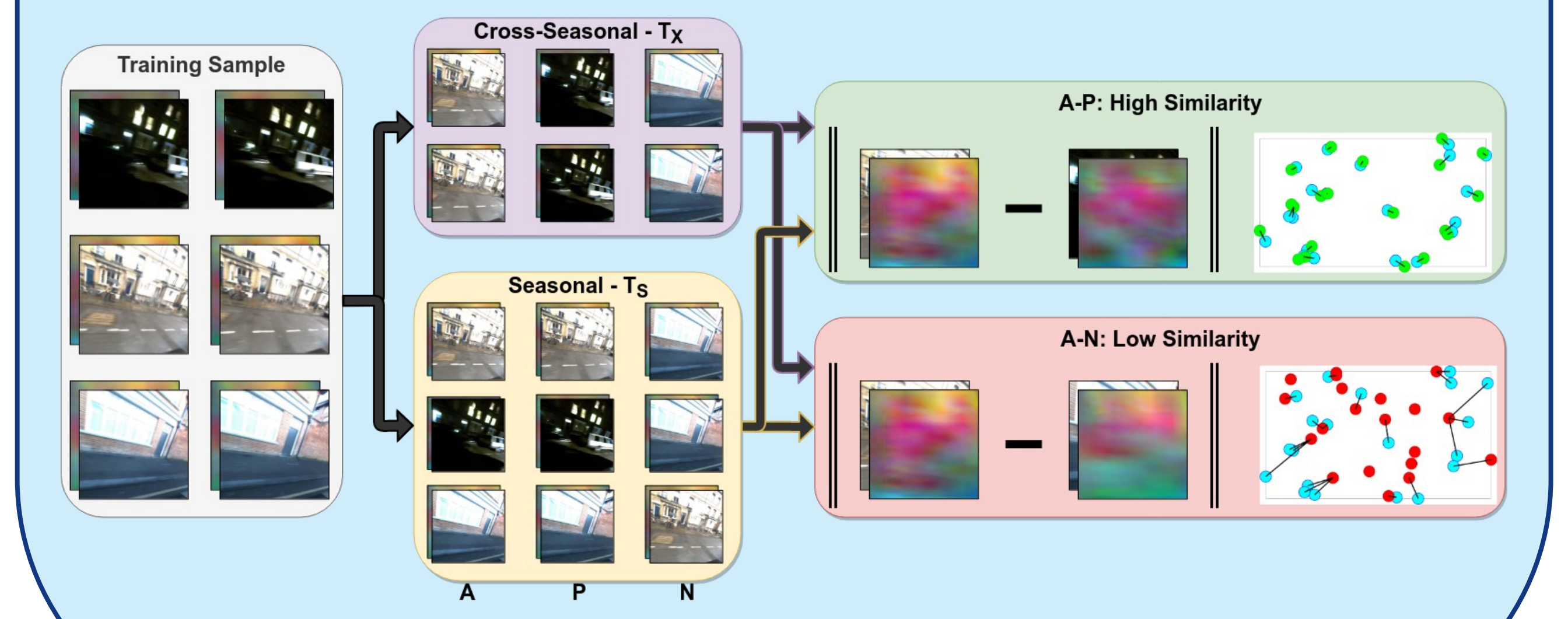
- Since we make use of **relational labels** between images (same location or not) we define a **triplet loss** based on this Contextual Similarity
- **Positive examples** should be marked as **highly similar**, so the triplet loss direction is modified accordingly

$$l(T) = \max(CX(\boldsymbol{F}_A, \boldsymbol{F}_N) - CX(\boldsymbol{F}_A, \boldsymbol{F}_P) + m, 0)$$

- We make use of both **seasonal and cross-seasonal triplets**

$$L = \frac{1}{N_X} \sum_{T_X} l(T_X) + \frac{\alpha}{N_S} \sum_{T_S} l(T_S)$$

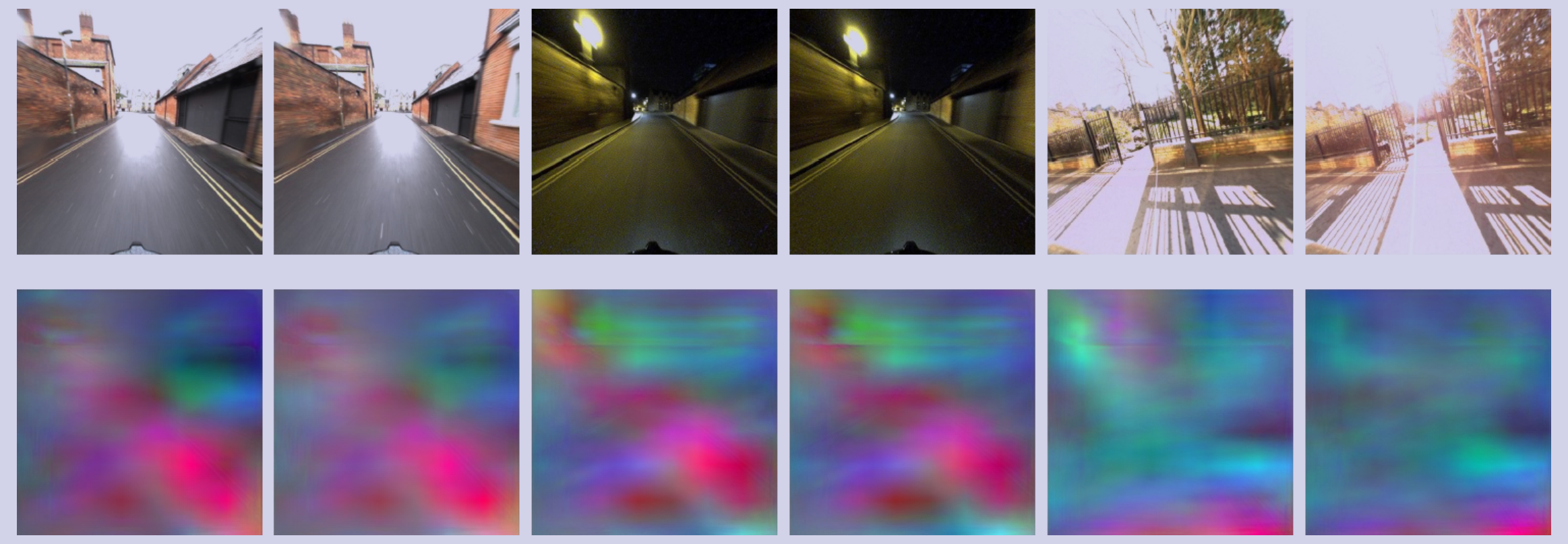
- α is a **balancing weight** in the range [0, 1], controlling the contribution of **same-season** triplets



7 - Evaluation

7.1 - Feature Visualization

- **Compact** feature representation restricted to **10-D**
- Features can be visualized by projecting them onto the **RGB cube via PCA**
- This visualization shows how despite **drastic appearance changes** we can still correctly identify the corresponding positive pair



Sample triplet visualization

7.2 - Cross-Seasonal AUC

- Akin to **image retrieval**
- **AUC** when **classifying** pair of images as corresponding to the **same location or not**
- Each **new location** in RobotCar Seasons as "**true positive**"
- The **contextual similarity** improves performance even in traditional methods, e.g. **ORB, (Root)SIFT**

Features	Seasonal AUC	Cross-season AUC
SIFT [28]	80.78	46.79
RootSIFT [1]	**97.15**	59.75
ORB [43]	96.60	**66.99**
SIFT + CX	94.42	64.58
RootSIFT + CX	95.55	68.36
ORB + CX	96.26	70.54
VGG [49] + CX	99.05	73.03
NC-Net [41] + CX	97.58	74.03
D2-Net [10] + CX	98.70	74.96
SAND [50] + CX	**99.74**	74.86
NetVLAD [2] + CX	99.41	**77.57**
DVF - $\alpha = 0$	99.30	93.82
DVF - $\alpha = 0.2$	**99.82**	**96.56**
DVF - $\alpha = 0.4$	99.59	91.37
DVF - $\alpha = 0.6$	99.76	93.46
DVF - $\alpha = 0.8$	99.52	94.12
DVF - $\alpha = 1$	99.47	92.94

Source Season \ Target Season	Dawn	Dusk	Night	N-R	O-S	O-W	Rain	Snow	Sun
Dawn	99.88	99.81	94.14	97.21	96.27	99.3	99.81	99.68	97.54
Dusk	99.81	99.99	89.88	94.58	98.87	99.21	99.89	97.99	95.22
Night	94.14	89.88	99.83	98.74	88.58	95.42	91	95.7	91.86
N-R	97.21	94.58	98.74	99.9	96.09	93.21	97.49	97.92	94.49
O-S	96.27	98.87	88.58	96.09	100	99.02	98.73	95.61	95.17
O-W	99.3	99.21	95.42	93.21	99.02	99.81	99.5	98.65	96.27
Rain	99.81	99.89	91	97.49	98.73	99.5	99.85	98.99	97.01
Snow	99.68	97.99	95.7	97.92	95.61	98.65	98.99	99.91	97.24
Sun	97.54	95.22	91.86	94.49	95.17	96.27	97.01	97.24	99.24

DejaVu (proposed)

Source Season \ Target Season	Dawn	Dusk	Night	N-R	O-S	O-W	Rain	Snow	Sun
Dawn	99.35	84.79	68.09	59.47	90.78	93.14	90.67	91.67	79.59
Dusk	84.79	99.92	65.92	62.51	84.01	93	97.95	90.63	76.43
Night	68.09	65.92	99.68	63.56	62.52	56.95	65.75	62.04	72.86
N-R	59.47	62.51	63.56	99.98	62.03	56.94	71.04	60.47	57.61
O-S	90.78	84.01	62.52	62.03	99.07	93.61	89.01	87.4	81.01
O-W	93.14	93	56.95	56.94	93.61	99.78	95.83	94.7	80.07
Rain	90.67	97.95	65.75	71.04	89.01	95.83	99.77	93.94	74.76
Snow	91.67	90.63	62.04	60.47	87.4	94.7	93.94	98.94	81.66
Sun	79.59	76.43	72.86	57.61	81.01	80.07	74.76	81.66	98.19

NetVLAD

7 - Evaluation cont'd

7.3 - Sparse Feature Matching

- Despite being trained **without spatial constraints**, features can be used in **sparse matching**
- Harris corner detector + NN matching + RANSAC
- **Improved matching** in cross-seasonal pairs

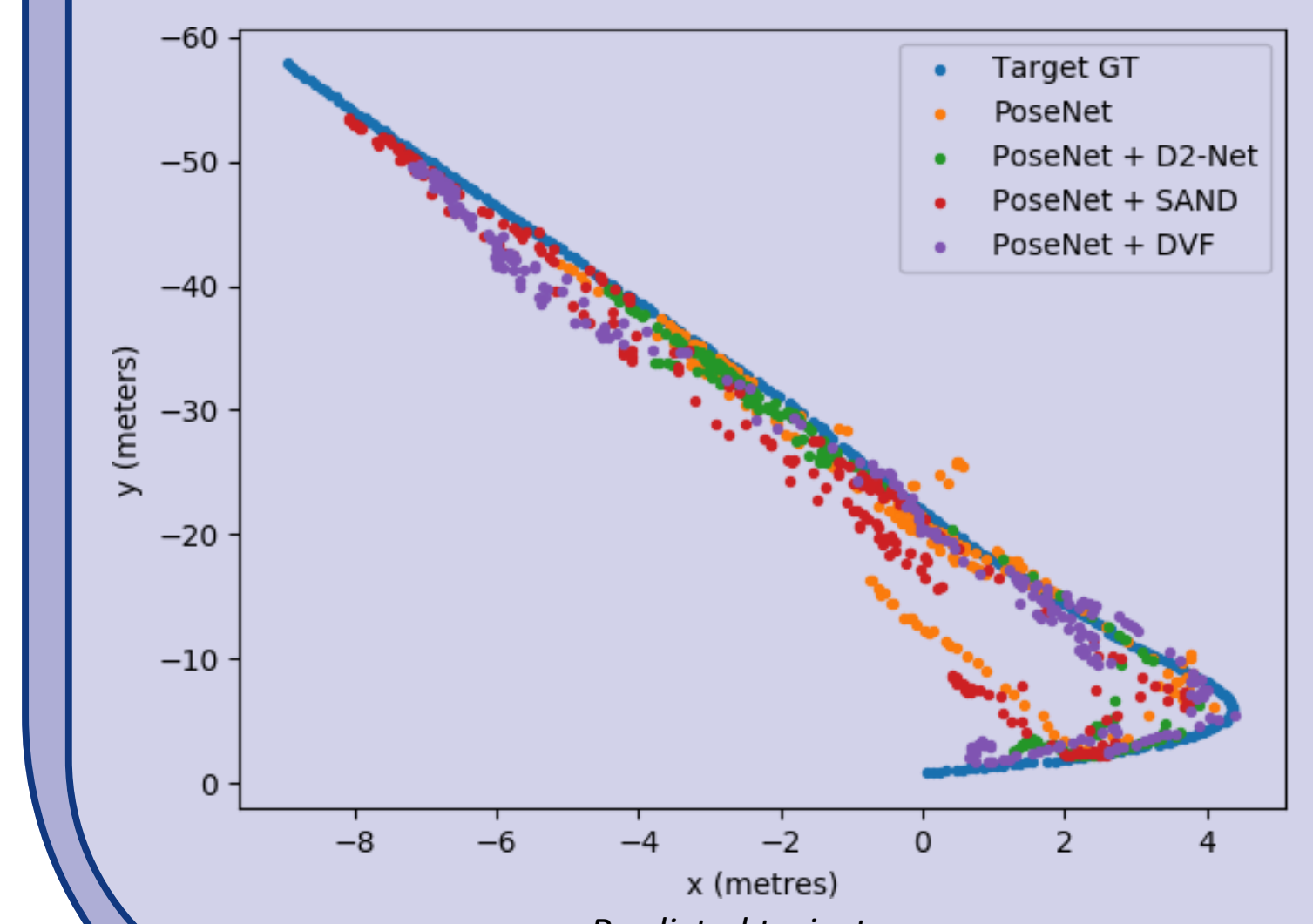
Deja-Vu (proposed)

D2-Net

SAND

7.4 - Cross-Seasonal Relocalisation

- Makes use of the full **dense features**
- **PoseNet** is **trained on snowy** data and **evaluated on overcast-sun** data
- The **input image** to PoseNet is **replaced with** its **dense feature** representation



Predicted trajectory

Method	**P** (m)	**R** (deg/m)
PoseNet [22]	10.3459	0.0170
D2-Net [10]	11.1858	**0.0029**
SAND [50]	7.3386	0.0045
DVF - $\alpha = 0$	**5.5759**	0.0050
DVF - $\alpha = 1$	7.2076	0.0036

Pose and Rotation errors

8 - Conclusions

- We introduce **Deja-Vu**, a framework for **weakly supervised cross-seasonal** feature learning
- This is one of the only approaches capable of learning dense features from a **holistic similarity metric**
- Despite this, the learnt features can still be used in **sparse matching** tasks
- Future work could include **incorporating spatial constraints** into the proposed loss

This work was funded by the EPSRC under grant agreement (EP/R512217/1).
We would also like to thank NVIDIA Corporation for their Titan Xp GPU grant.

EPSRC Engineering and Physical Sciences Research Council