GAN-based Image Colourisation with Feature Reconstruction Loss

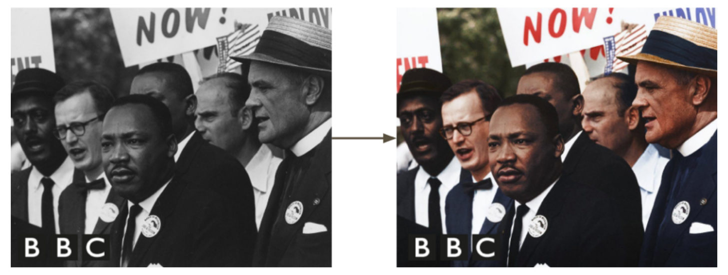
WiCV @ CVPR — Women in Computer Vision — VIRTUAL JUNE 19-25

Laia Tarrés — MsC Student @ UPCTech
Xavier Giró — Teacher & Researcher @ UPCTech
Marta Mrak — Researcher @ BBC R&D
Marc Górriz — Researcher @ BBC R&D

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona
telecos BCN
BBC

RELATED WORK

Our work focuses in the basic image enhancement of **colourisation**, adding colour to grayscale media.



Colorful Image Colourisation:

- Only use **colourisation loss** (L1).
- Produces desaturated results.

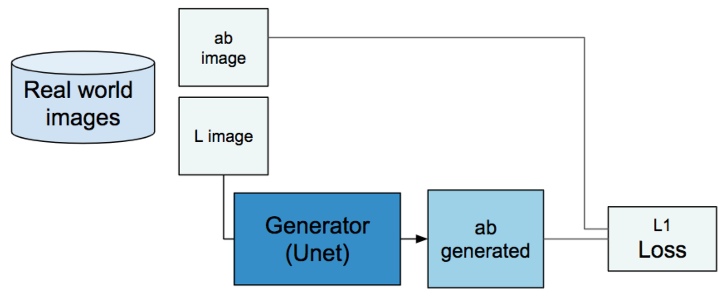
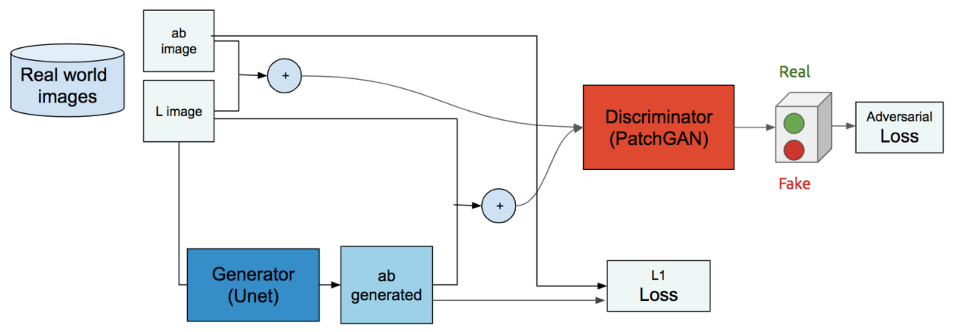


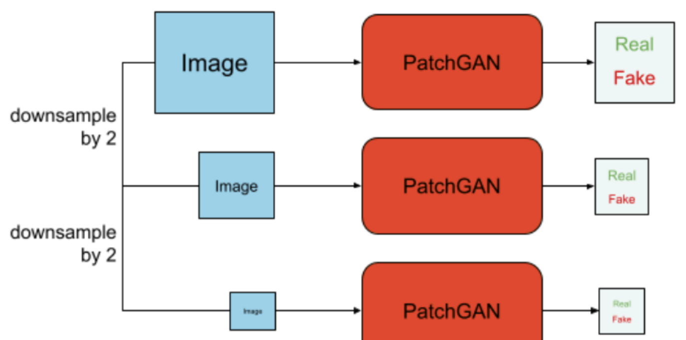
Image-to-image translation with conditional adversarial networks (pix2pix):

- Adds **adversarial loss**.
- Unet architecture for the generator.
- PatchGAN architecture for the discriminator.



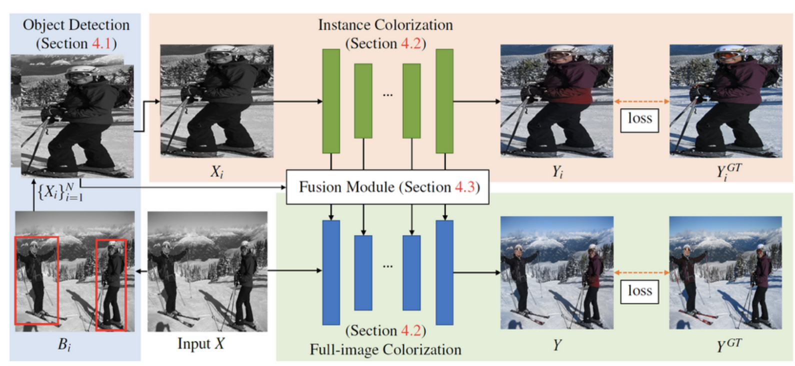
End-to-End Conditional GAN-based Architectures for Image Colourisation (Górriz et al.):

- Adds a combination of batch and instance normalization (**IBN**).
- Adds **spectral normalization** as a regularizaton step.
- Adds a **multi-scale discriminator**.



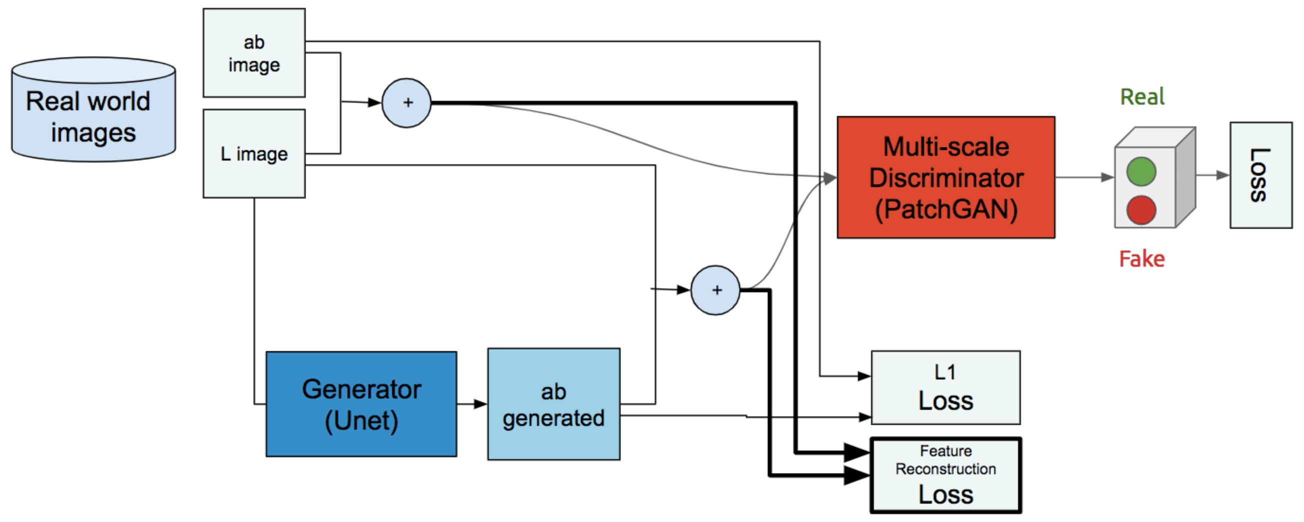
Instance-aware image colorization:

- Proposes **detecting** the objects to colourise them individually.
- Adds a **fusion module** to colourise the whole image.

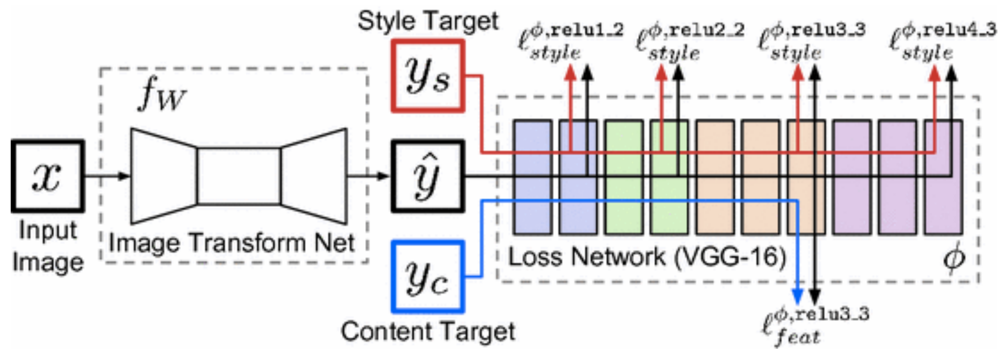


PROPOSED METHOD

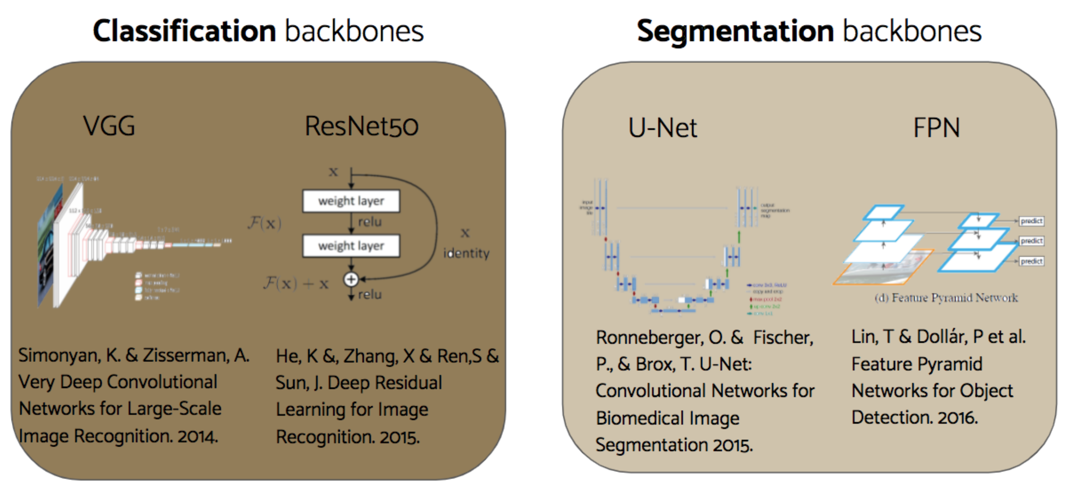
We introduce the **feature reconstruction loss** to train the network.



The feature reconstruction loss [1] is computed as the squared and normalized euclidean distance between activations in a punctual layer of the network for the output and target image.

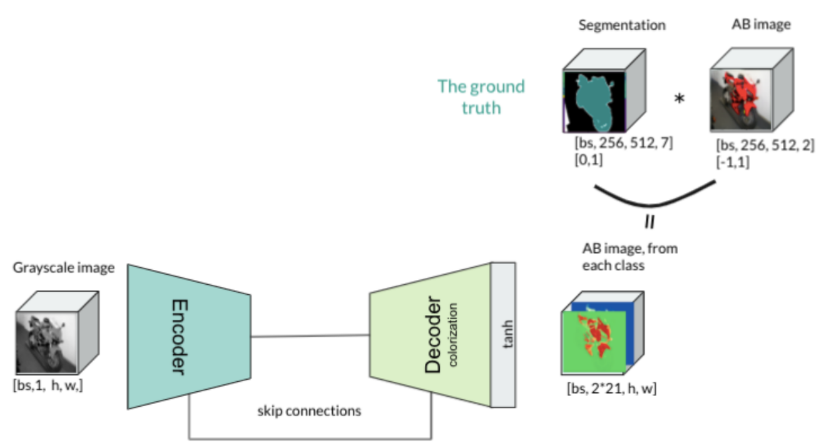


For our implementation, we experimented with different **backbones** (networks from which we compute the activations):

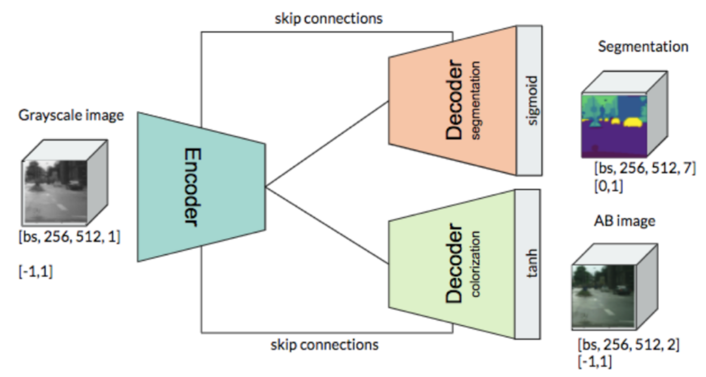


In our extended abtract, we also propose two different approaches to introduce a **segmentation** loss:

With a **shared** decoder:



With **separate** decoders:

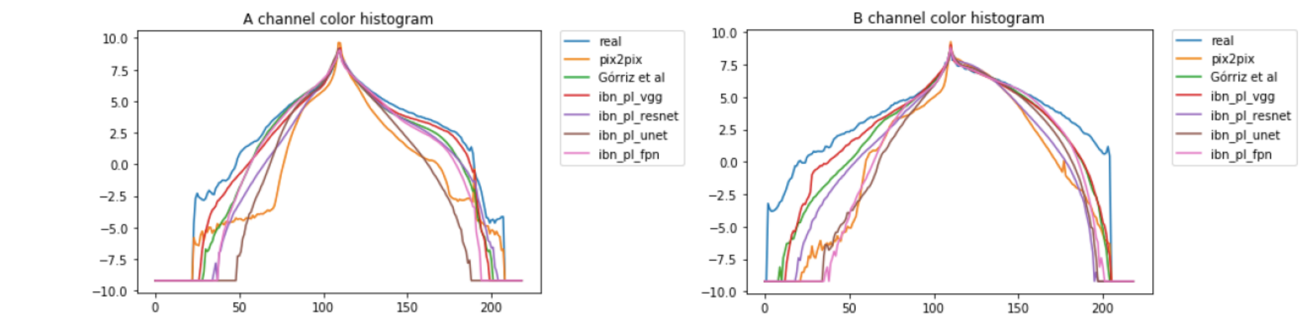


EXPERIMENTAL RESULTS

We provide a table with the **quantitative results**. We have chosen pixel-wise metrics: Peak Signal-to-Noise Ratio (**PSNR**), as well as histogram divergences: Kullback Liebler divergence (**KL**) and Jensson Shanon divergence (**JS**).

Models	Backbone	JS divergence ↓		KL divergence ↓		PSNR ↑
		a	b	a	b	
pix2pix [4]	-	0.13	0.13	79.52	82.94	**26.70**
Górriz et al. [7]	-	**0.06**	0.06	59.51	22.28	25.14
Our model	VGG	0.09	**0.05**	**50.33**	**20.15**	25.13
	ResNet	0.12	0.13	94.82	109.70	25.23
	U-Net	0.23	0.19	282.60	205.80	25.19
	FPN	0.15	0.19	139	197.90	25.24

We also provide the **histograms** for the colour channels. The histogram for the real images is the widest, which denotes more vivid colours in the resulting images. It is followed by **our model with the VGG backbone**.



We performed a **perceptual realism study**. 35 Non-expert participants were shown images coloured with: real, pix2pix, Górriz et al and our best-performing model. For each image shown, the participant has to indicate if the image has real or generated colours. The naturalness metric corresponds to the % of pictures noted as real from each model.

Model	Naturalness
Ground Truth	0.87
pix2pix [4]	0.53
Górriz et al. [7]	0.26
Ours with VGG	0.38

We also show some results for the baselines and our best performing model:

[1] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." ECCV 2016.

Acknowledgements: Generalitat de Catalunya; Gobierno de España, Ministerio de Economía, Industria y Competitividad