End-to-End Conditional GAN-based Architectures for Image Colourisation

Marc Gorriz Blanch [1], Marta Mrak [1], Alan F. Smeaton [2], Noel E. O'Connor [2]

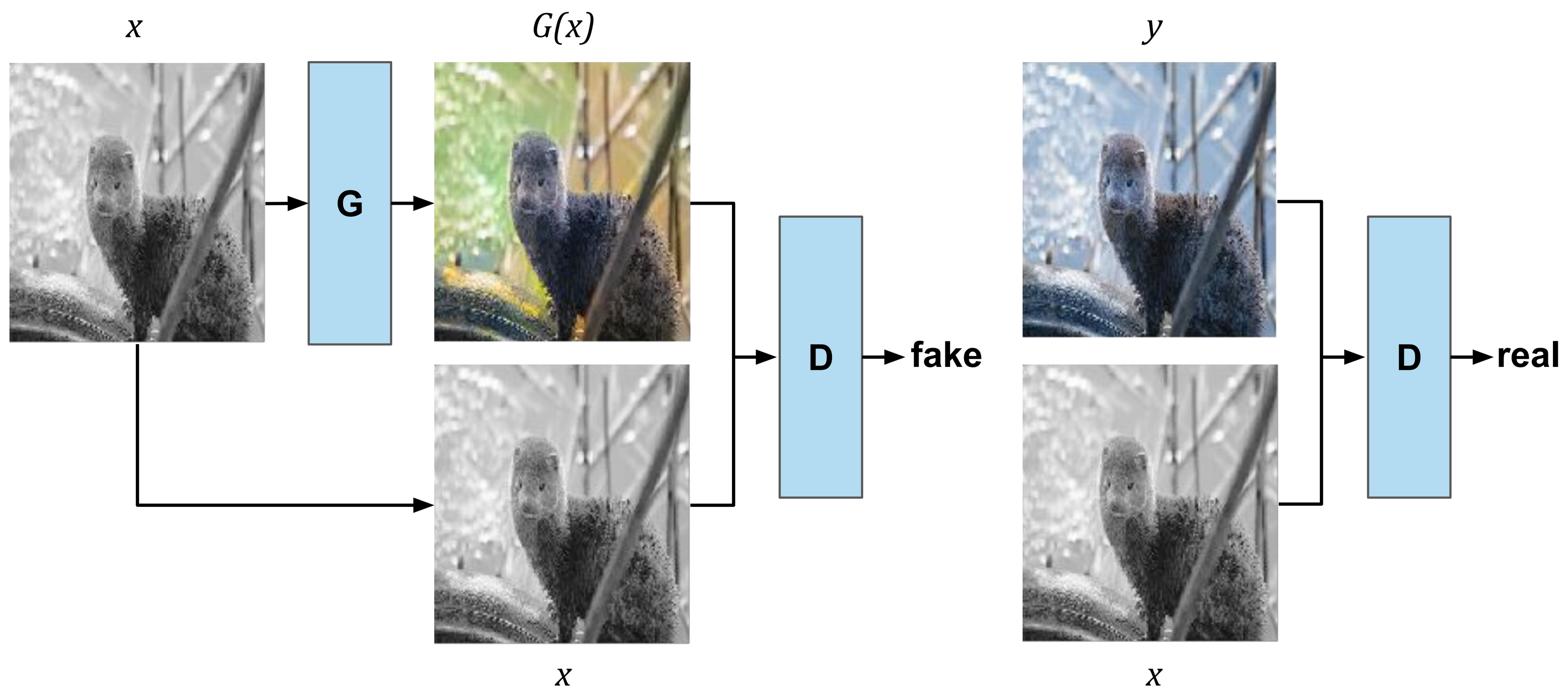
BBC R&D DCU

https://bbc.in/2mrNqyk

Motivation

In this work recent advances in **conditional adversarial networks** are investigated to develop an **end-to-end architecture** based on **Convolutional Neural Networks (CNNs)** to directly map realistic colours to an input greyscale image. Observing that existing **colourisation methods** sometimes exhibit a **lack of colourfulness**, this paper proposes a method to improve colourisation results. In particular, the method uses **Generative Adversarial Neural Networks (GANs)** and focuses on **improvement of training stability** to enable better generalisation in large multi-class image datasets. The contributions of this work are the following:

- A novel **generator-discriminator setting** which **adapts the IBN paradigm** to an **encoder-decoder architecture**, enabling generalisation of the content's style changes while encouraging stabilisation during GAN training.
- The use of **Spectral Normalisation** for **improving the generalisation** of the adversarial colourisation and **preventing training instability**.
- The use of **multi-scale discriminators** to achieve an **improved colour generation in small areas and local details** and a **boosted colourfulness**.



Improved Conditional GAN architecture

We improve the capabilities of adversarial models for image colourisation by adapting an **Instance-Batch Normalisation (IBN) convolutional architecture** to an **end-to-end conditional GAN.**

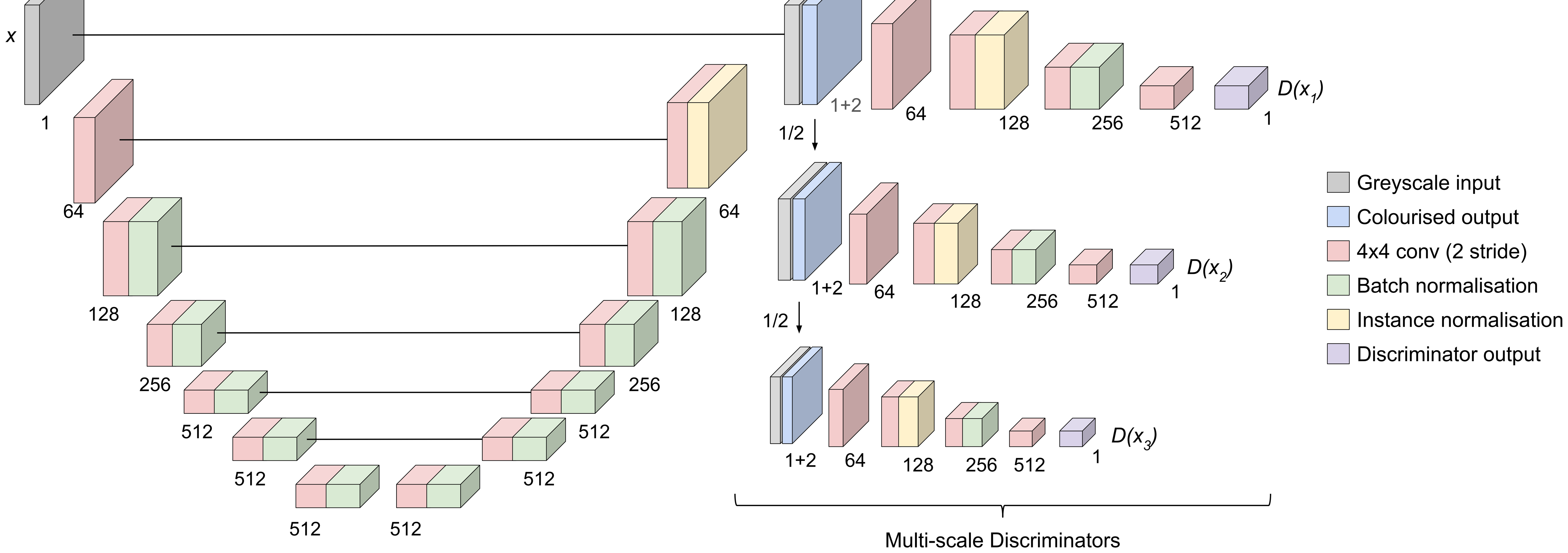
- **Instance Normalisation (IN)** uses the statistics of an individual sample to capture style changes invariant to style information.
- **Batch normalisation (BN)** reduces the covariance shift within the whole batch to stabilise the GAN learning and to prevent the mode collapse due to poor initialisation.

We apply **Spectral Normalisation (SN)** to regularise the network weights and to prevent unbounded gradients in the discriminator when using a sigmoid activation. While this technique prevents instability during training, helps to generalise the style changes of large multi-class datasets leading to more colourful results.

Multi-scale discrimination (MD): fixed architecture + variable receptive field.

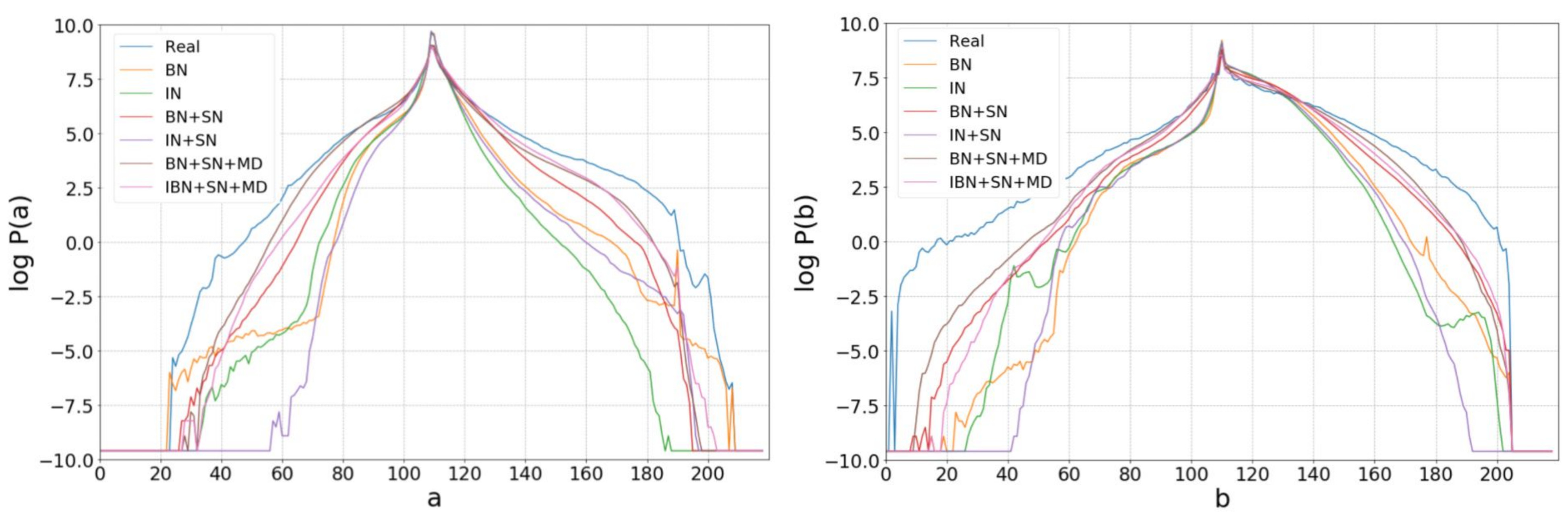
$$V'(G, D) = \sum_{n=0}^{N-1} V(G, D_n)$$

where $\{X_n, (G(X))_n\} \in \mathbb{R}^{M \times H_n \times W_n \times C}$ are the D_n inputs

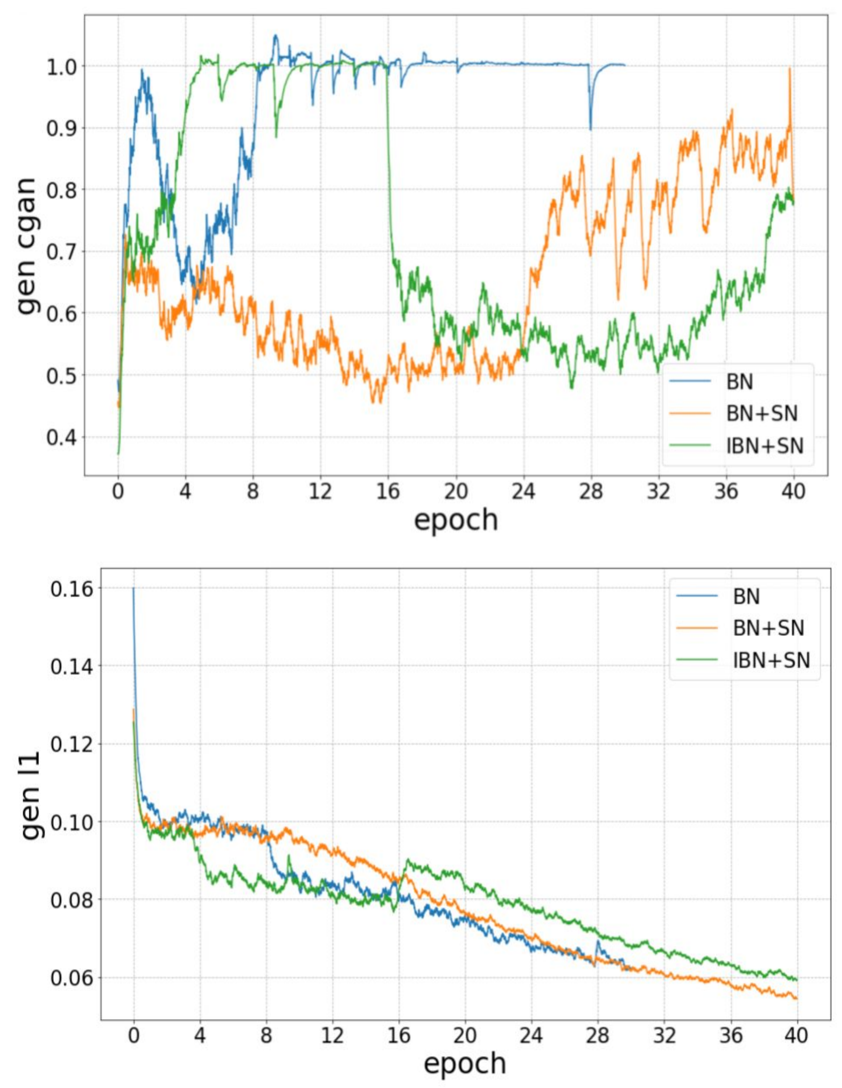


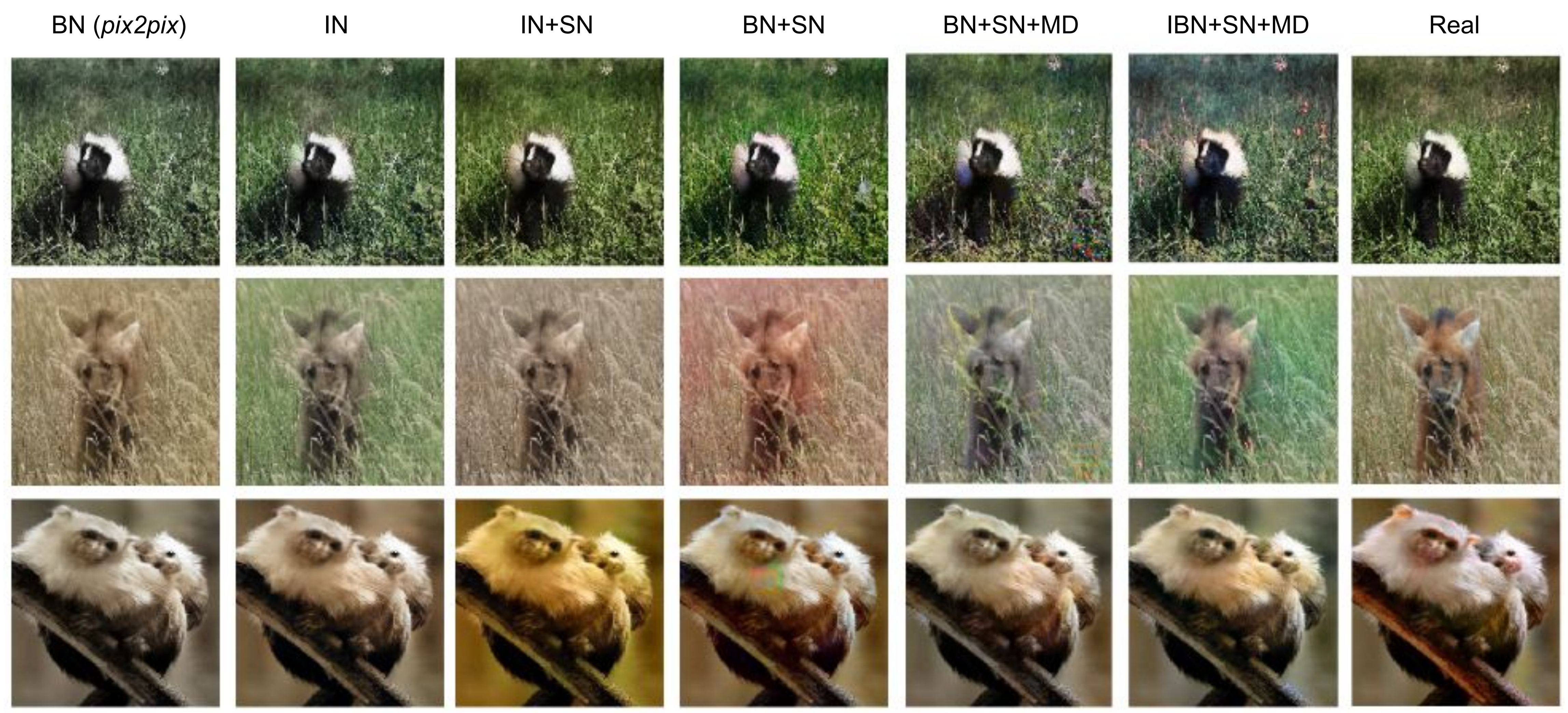
Colourfulness Evaluation

A common issue is the **desaturated effect**, characterised by **low absolute values in the colour *ab* channels**, when an **automatic colourisation model** is trained on **large databases of natural images**. Observe the **colourfulness boosting** of the proposed configurations over the baseline *pix2pix* model (BN).



Experimental Results





(1) BBC Research and Development. London, UK.
(2) Dublin City University. Dublin, Ireland.

jolt

The work described in this paper has been conducted within the project JOLT. This project is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No 765140.