InstanceNorm

A Preprint

1 Overview

Instance Normalization (IN) is a normalization technique introduced to improve the quality of image stylization networks. It replaces Batch Normalization (BN) in generator architectures and is applied both during training and inference.

The key insight is that stylization quality benefits from removing instance-specific contrast information from the content image.

2 Mathematical Definition

Let the input be a 4D tensor:

$$x \in \mathbb{R}^{T \times C \times H \times W}$$

where T is batch size, C is number of channels, $H \times W$ are spatial dimensions.

For each image t and channel i:

$$\mu_{t,i} = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} x_{t,i,h,w}, \quad \sigma_{t,i}^2 = \frac{1}{HW} \sum_{h,w} (x_{t,i,h,w} - \mu_{t,i})^2$$

Normalization step:

$$\hat{x}_{t,i,h,w} = \frac{x_{t,i,h,w} - \mu_{t,i}}{\sqrt{\sigma_{t,i}^2 + \varepsilon}}$$

Affine transformation (learnable):

$$y_{t,i,h,w} = \gamma_i \cdot \hat{x}_{t,i,h,w} + \beta_i$$

3 Why InstanceNorm Works Better in Stylization

- Removes contrast variation: Stylization should reflect the contrast of the style image, not the content image. IN removes contrast cues from the content image, which simplifies learning.
- Sample-independent: Unlike BatchNorm, IN does not rely on batch statistics, making it more stable for small batch sizes and test-time inference.
- Focus on spatial consistency: Since IN normalizes per-instance per-channel, it enhances the model's ability to reproduce coherent textures and style patterns within each image.
- Avoids batch-induced artifacts: BN introduces variability depending on the composition of the batch. IN avoids this by operating independently on each image.
- Improves generalization for generators: It removes nuisance variation and helps the network learn style representations that transfer more cleanly.

4 Comparison to BatchNorm

- BatchNorm: Computes mean/variance across the batch and spatial dimensions.

$$\mu_i = \frac{1}{THW} \sum_{t,h,w} x_{t,i,h,w}$$

- InstanceNorm: Computes statistics only per image and channel.

$$\mu_{t,i} = \frac{1}{HW} \sum_{h,w} x_{t,i,h,w}$$

5 Conclusion

Instance Normalization is a lightweight architectural change with a large impact on visual quality in image generation tasks. It improves stylization networks by discarding unwanted content-specific contrast and enabling consistent texture rendering.