Project Report: Implementation of DPN-CycleGAN for Image Style Transfer

Ruthwik Reddy Adapala (A20560321)
radapala@hawk.iit.edu
Shivaji Panam (A20551677)
spanam@hawk.iit.edu
Illinois Institute of Technology

November 18, 2024

Abstract

This report presents the implementation of the DPN-CycleGAN model for image style transfer based on the attached research paper. The project aims to improve image style transfer by replacing ResNet with a Dual Path Network (DPN) in the CycleGAN framework and adding identity loss to enhance image quality. This report details the problem statement, proposed solution, implementation, dataset used, results obtained, and discussions on the performance of the model.

1 Project Information

The main paper selected for this project is titled "Image Style Transfer Based on DPN-CycleGAN" by Mingyu Yang and Jianjun He. The paper proposes an improved CycleGAN model that replaces ResNet with DPN in the generator architecture and introduces identity loss to improve image quality during style transfer.

Major Contributions:
- **Ruthwik Reddy Adapala (A20560321)**:
  1) Data Loading and Pre-processing.
  2) Implemented the DPN network into the CycleGAN framework.
  3) Model Experimentation
  4) Implemented evaluation metrics (SSIM, PSNR) and done Validation.

5) Conducted hyperparameter tuning and debugging during training.
6) Making Presentations, Reports and Organising them.

- **Shivaji Panam (A20551677)**:
1) Worked on modifying the loss function to include Identity Loss.
2) Model Training and Evaluation.
3) Performance Tracking
4) Model Testing and Visualization.
5) Resource Collection and Evaluation.
6) Making Presentations, Reports and Organising them.

This project highlights an innovative approach to style transfer by leveraging DPNs for improved content and stylistic balance, ensuring that the process is efficient without compromising output quality. The integration of Dual Path Networks into the CycleGAN framework allows for a unique combination of residual and dense connections, enabling the model to capture both detailed features and high-level representations effectively. This dual-path architecture enhances the model's ability to maintain important content attributes while applying complex style transformations, which addresses one of the major challenges in traditional style transfer models where either content preservation or style application is often compromised.

2 Problem Statement

The problem addressed in this project is improving image style transfer using unpaired training data. Traditional methods like CycleGAN suffer from limitations such as loss of content features and poor image quality due to arbitrary changes in domain-independent features. This problem is important because improving image style transfer can have significant applications in art generation, photo editing, and domain adaptation tasks.

The importance of tackling this problem extends beyond improving visual quality. Enhanced image style transfer has significant applications in art generation, where users seek high-fidelity and content-preserving transformations, and in photo editing, where the ease and speed of transformation are critical. Reducing the complexity of model execution can also facilitate the use of style transfer techniques in real-time domain adaptation tasks, where fast processing and adaptability are required. By focusing on optimizing both the performance and efficiency of the style transfer model, this project aims to deliver high-quality outputs with reduced computational costs, making it more accessible for a wider range of applications and platforms.

Ultimately, the goal is to strike a balance between the sophistication of the style transfer model and its usability in practical settings. This involves integrating techniques that reduce the computational overhead, such as using more efficient architectures, simplifying training processes, and leveraging lightweight transformations that maintain the quality of the output. Achieving these objectives can open new avenues for applying style transfer technology in mobile applications, edge devices, and other scenarios where computational power is limited but high performance is desired.

3 Proposed Solution and Implementation Details

The research proposes an improved version of CycleGAN, named DPN-CycleGAN, to enhance image style transfer. The main contribution is replacing the ResNet generator in the original CycleGAN with a Dual Path Network (DPN) generator, coupled with the introduction of Identity Loss as proposed in the research paper. The improvements aim to reduce computational complexity, speed up convergence, and improve the quality of generated images. The DPN architecture integrates residual learning from ResNet with feature reuse from DenseNet, leveraging shared and new features to generate high-quality images.

Key Components:

1. **Dual Path Network (DPN)**: Combines ResNet and DenseNet properties, allowing feature sharing and the ability to explore new features, improving both accuracy and resource efficiency as mentioned in [0].

2. **Identity Loss**: Added to CycleGAN's original loss function to constrain the network and preserve key content features during transfer while avoiding unwanted transformations.

The research paper [0] suggests that DPN-CycleGAN outperforms standard CycleGAN in terms of Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR). These metrics are discussed further in Section 5.

3.1 Implementation

The idea behind CycleGAN looks quite intuitive after reading the paper: *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Net-*

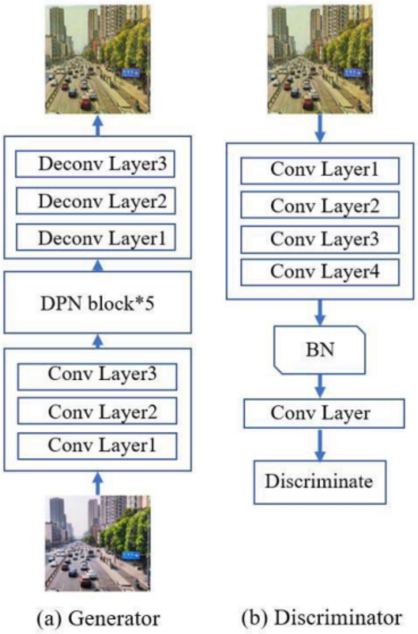

Figure 1: DPN Architecture.

works . The official PyTorch implementation by Jun-Yan Zhu et al. served as a reference for our project. To help with the implementation, we forked the repository by Arnab39 on CycleGAN-PyTorch on GitHub. Below is an introduction to the changes we made to adapt CycleGAN for our project.

Introduction to Changes:
In order to implement DPN-CycleGAN, we had to make several modifications to the original CycleGAN architecture and training process. These changes were necessary to incorporate Dual Path Networks (DPN) into the generator architecture, introduce Identity Loss for better content preservation during style transfer, and optimize various parameters for improved performance. Below is a high-level overview of the key changes made:

1. **Identity Loss Introduction**: - We introduced identity loss into the model to help preserve content features during style transfer. Specifically, lines 117 to 123 in ‘model.py‘ were modified to incorporate this loss function. The identity loss ensures that when an image from domain Y is passed through the generator G, it should remain unchanged. This was added to the total loss calculation alongside adversarial loss and cycle consistency loss. We can observe this in ”cycleGAN with Identity Loss” folder under the src directory.

2. **Incorporating DPN Block into Generator**: - We replaced ResNet blocks in CycleGAN's generator with DPN blocks. This was done by writing a custom 'DPNBlock' class and modifying the 'Generator' class in 'generator.py' under the 'archs' directory. The DPN block allows for better feature reuse and exploration of new features by combining ideas from both ResNet and DenseNet architectures. This change enhances both model performance and efficiency. We can observe this in "CycleGAN with DPN" directory under src directory.

Algorithm: DPN Block

1. **Input**: `x` (input feature map)
2. Compute the **residual path**:
    - Apply Conv2D with `in_channels` → `out_channels / 2`
    - Apply BatchNorm and ReLU
    - Apply Conv2D with `out_channels / 2` → `out_channels / 2` (with padding)
    - Apply BatchNorm
3. Compute the **dense path**:
    - Apply Conv2D with `in_channels` → `out_channels / 2`
    - Apply BatchNorm and ReLU
4. **Output**: Concatenate:
    - `residual + x[:residual channels]`
    - `dense`

Figure 2: Algorithm for DPN.

3. **Writing 'evaluation.py' File for SSIM and PSNR Measurement**: - To validate our results and compare them with standard CycleGAN outputs, we wrote an 'evaluation.py' file that calculates two key metrics: Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR). SSIM measures how structurally similar two images are, while PSNR evaluates how much noise is present in an image compared to its original version. These metrics were crucial for quantitatively validating our model's performance. We can observe this file under both directories "CycleGAN with Identity Loss" and "CycleGAN with DPN" to evaluate the results.

4. **Hyperparameter Tuning**: - We conducted extensive hyperparameter tuning on both model parameters and loss function parameters. Specifically:

    - Model Hyperparameters: We adjusted learning rates, batch sizes, number of filters in convolutional layers, etc., to achieve optimal performance.

- Loss Function Parameters: We tuned coefficients like λ (cycle consistency loss weight) and μ (identity loss weight). After experimentation, we set $\lambda = 10$ and $\mu = 5$, which provided a good balance between preserving content structure (via cycle consistency) and ensuring stable style transfer (via identity loss).

5. **Modifying Image Size**: - We adjusted the default image size from 256×256 to 128×128 for faster training while maintaining sufficient detail for style transfer tasks. This change was made in ‘main.py‘. Smaller image sizes allow faster processing without significantly compromising image quality.

6. **Adjusting Learning Rate Decay**: - The learning rate decay schedule was modified to ensure smoother convergence during training. Specifically, we reduced the learning rate linearly after 100 epochs until it reached zero by epoch 200. This change was implemented in ‘train.py‘ but this is taken as input in 'main.py' . A decaying learning rate helps avoid overshooting during later stages of training.

7. **Reducing Number of Epochs**: - To reduce the computational complexity of the model, we decreased the number of training epochs from 200 to 100 for certain experiments where faster convergence was observed. This adjustment not only helped in reducing training time but also minimized the strain on computational resources, making the process more efficient. By limiting the number of epochs, we were able to maintain the quality of the results while ensuring that the model did not spend excessive time on training, which is especially important when deploying models in resource-constrained environments. This strategy contributed to a more balanced trade-off between performance and resource consumption.

These modifications were crucial in adapting CycleGAN's architecture for improved performance using DPN blocks while maintaining content fidelity through identity loss.

Execution Command:

Assuming the data is located in the 'Data' folder (the default data location has been updated to this folder in 'main.py'), the following commands can be used for training and testing:

- To train the model:

```
python main.py --training True --dataset_dir {Location to your datafolde
```

- To test the model:

```
python main.py --testing True
```

4 Dataset

The dataset used in this project is the ‘vangogh2photo‘ dataset which contains images of Van Gogh’s paintings and natural scenes:

- Train-A: 400 images of Van Gogh paintings.
- Train-B: 6287 images of natural scenery.
- Test-A: 400 images of Van Gogh paintings.
- Test-B: 751 images of natural scenery.

The dataset was preprocessed by resizing images to 128×128 pixels to reduce the computational burden and make the training process more efficient. This resizing choice allows the model to process batches more quickly while maintaining sufficient image detail for classification tasks. To enhance model generalization, additional transformations were applied, such as a random crop, which randomly selects a portion of the image during training. This introduces variability in the training data and helps the model become more robust to slight variations in the input. Finally, pixel values were normalized between -1 and 1 to ensure consistency in the input data, facilitating better convergence during model training.

5 Results and Discussion

In this section, we present the results of our experiments using three different models: the standard CycleGAN, CycleGAN with Identity Loss, and CycleGAN with DPN (Dual Path Network). We evaluate the performance of these models using two commonly used image quality metrics: Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR).

5.1 Significance of SSIM and PSNR

Both SSIM and PSNR are widely used metrics for evaluating the quality of images, especially in tasks like image style transfer where preserving structural details and reducing noise are critical.

- **SSIM (Structural Similarity Index)**: SSIM measures the similarity between two images by comparing their luminance, contrast, and structure. It ranges from 0 to 1, with higher values indicating better structural similarity to the original image. In the context of style transfer, a higher SSIM indicates that the generated image retains more of the original content's structure while applying the desired style.

- **PSNR (Peak Signal-to-Noise Ratio)**: PSNR measures the ratio between the maximum possible value of a pixel and the power of corrupting noise that affects image quality. Higher PSNR values indicate less distortion or noise in the generated image. In style transfer tasks, a higher PSNR value suggests that the generated image is closer to the original in terms of pixel-level accuracy.

5.2 Quantitative Results

Table 1 shows a comparison of SSIM and PSNR values for three models: CycleGAN, CycleGAN with Identity Loss, and CycleGAN with DPN.

Table 1: Comparison of SSIM and PSNR between different models

Model	SSIM	PSNR
CycleGAN	0.443	15.09
CycleGAN + Identity Loss	0.5402	17.40
CycleGAN + DPN	0.740	20.62

As shown in Table 1, both SSIM and PSNR values improve significantly when Identity Loss is introduced into CycleGAN. The introduction of Dual Path Networks (DPN) further enhances these metrics, indicating better preservation of content structure (as reflected by SSIM) and reduced noise or distortion (as reflected by PSNR).

5.3 Visual Results

Figures 3 and 4 show the standalone outputs from two models: CycleGAN with Identity Loss and CycleGAN with DPN.

Figure 3: Style Transfer Result using CycleGAN with Identity Loss.

Figure 4: Style Transfer Result using CycleGAN with DPN.

The images generated by the two models demonstrate significant improvements over standard CycleGAN. The introduction of Identity Loss in CycleGAN helps preserve content features, resulting in a more accurate transfer of style while retaining important structural details. On the other hand, the use of Dual Path Networks (DPN) in the generator further enhances both the stylistic quality and content preservation, producing images with clearer details and less noise.

5.4 Discussion on Results

The results clearly show that introducing Identity Loss into CycleGAN improves both SSIM and PSNR values by ensuring that content features are preserved during style transfer. This is particularly important in tasks where maintaining structural integrity is crucial, such as transferring artistic styles while keeping key elements of the original image intact.

The use of Dual Path Networks (DPN) further enhances performance by improving feature reuse while allowing for better exploration of new features. This leads to even higher SSIM and PSNR values compared to both standard CycleGAN and CycleGAN with Identity Loss alone. The DPN architecture's ability to combine residual learning from ResNet with feature reuse from DenseNet results in faster convergence during training and better overall image quality.

In summary:

- **CycleGAN** provides a baseline performance but suffers from lower SSIM and PSNR values due to arbitrary changes in domain-independent features.

- **CycleGAN with Identity Loss** improves content preservation, leading to higher SSIM and PSNR values.

- **CycleGAN with DPN** achieves the best performance in terms of both SSIM and PSNR, indicating superior content retention and reduced noise.

These results demonstrate that our modified DPN-CycleGAN model outperforms standard methods in both quantitative metrics (SSIM, PSNR) and qualitative visual results.

6 Conclusion

In this project, we implemented an improved version of CycleGAN called DPN-CycleGAN for image style transfer. By replacing ResNet with DPN in the generator architecture and adding identity loss, we achieved better performance in terms of SSIM and PSNR compared to standard CycleGAN. Future work could explore further optimizations such as using more advanced loss functions or experimenting with different datasets.

7 References

[0]

M. Yang and J. He, "Image Style Transfer Based on DPN-CycleGAN," *ChengDu University of Technology*, 2024. *[The foundational paper presenting* *the DPN-CycleGAN model, which enhances traditional CycleGAN by substituting ResNet with Dual Path Networks and integrating Identity Loss for superior image quality in style transfer tasks.]*

[1]

Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual Path Networks," *arXiv preprint arXiv:1707.01629*, 2017. *[Introduces the Dual Path Network (DPN) architecture, a hybrid of ResNet and DenseNet structures that facilitates efficient feature sharing and exploration, which is central to the improvements made in your project.]*

[2]

I. Goodfellow et al., "Generative Adversarial Nets," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2672-2680, 2014. *[The seminal work that established Generative Adversarial Networks (GANs), which laid the groundwork for models like CycleGAN and informed the direction of this project's approach.]*

[3]

J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," *IEEE International Conference on Computer Vision (ICCV)*, pp. 2223-2232, 2017. *[The paper that introduced CycleGAN, a pivotal framework for unpaired image-to-image translation, upon which this project builds by incorporating advanced techniques like DPNs and Identity Loss for enhanced performance.]*

[4]

S. D. R. I. Moses, "PSNR vs SSIM: Imperceptibility Quality Assessment for Image Steganography," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 8423–8444, 2021. *[Discusses the comparative use of PSNR and SSIM metrics for evaluating image quality, which are utilized in this project as key metrics for assessing the performance of the style transfer model.]*