Image Outpainting for Crop Dataset

Course: Deep Learning

Harsh Mishra (U20220039)
Anshika Singh (U20220015)
Rohit Singh (U20220074)
Govind (U20220037)

Plaksha University

Objective

The objective of this project is to tackle the task of image outpainting for a crop (rice) dataset by using deep learning methods. Image outpainting or extrapolation is the technique of generating an image beyond its boundaries (given a certain portion of the image, we make a complete image - generating what is missing from a small portion).

The goal of this project is to create realistic crop images that are not limited by the image data, using a GAN-based technique. Our goal is to create a model that can provide high-quality, coherent picture extrapolations for use in agriculture.

Dataset

We use a crop (rice) dataset composed of diverse images captured in real agricultural environments. The dataset was collected from different cities in Pakistan such as Kandhkot, Shikarpur, Sukkur, Moro, and Kashmore.

Figure 1: Screenshot of the crop dataset showing diversity in size, color, and texture.

This dataset provides visual data that can be used for agricultural tasks like disease prediction and weed detection, which is only possible if we have good images with consistent size and resolution. Image outpainting helps in achieving this.

The dataset is composed of 1007 images:

- Healthy crop images: 501
- Unhealthy crop images: 506

The images differ in terms of:

- Size
- Color
- Texture

The healthy images are bright green and have well-formed grain, whereas unhealthy crops are discolored and partially damaged. Using this visual inspection, the labeling was done.

The dataset was taken from Kaggle, and a DSLR (megapixel camera) was used to capture the images. The images are of varying dimensions, reflecting the diversity in capturing conditions. This dataset will be used for image outpainting as part of a larger project to restore and enhance incomplete or low-quality images. Certain images (such as "Healthy 1," "Unhealthy 2," and "Unhealthy 5") have defects or missing portions, making them unsuitable for training models for disease prediction or weed detection. Image outpainting will fill these gaps to improve their utility in machine learning tasks.

Potential Preprocessing Steps

- **Normalization**: To bring consistent pixel intensity across images.
- **Resizing**: Images are of varying sizes, resizing them to a uniform resolution.
- **Enhancement**: To improve visibility.
- **Contrast Adjustment**: To bring visual comparison across images.
- **Flipping and Saturation adjustment**.

Dataset Splitting

- Training: 80%
- Validation: 10%
- Test: 10%

Training

During training, the model receives an incomplete or masked image as input and learns to predict the surrounding regions and give it as output.

GitHub Link

G. Govind, "Weed Detection," GitHub repository. Available: `https://github.com/govindvgd/Weed-Detection`

Introduction

Compared to the art of inpainting, which focuses on rebuilding missing areas within an image, Image Outpainting is a relatively unexplored field. The peculiar challenge of outpainting is that it requires:

- Preserving the original input's and the created regions' spatial consistency.
- Producing a large, high-quality image with minimal surrounding data.

Agriculture can benefit greatly from this work, especially when it comes to expanding crop photos for improved analysis, prediction, and visualisation.

Importance of the Problem

Image outpainting involves predicting and creating visually coherent extensions of a picture beyond its initial bounds. This is especially useful in the context of crop datasets, where photos frequently show only a portion of a crop or plant. Extending these photos can improve visual analysis for resource management, crop monitoring, and agricultural research.

Reconstructing obscured or partial images using outpainting is essential for enhancing the accuracy of crop production forecasts, disease detection, and general farm management. Image outpainting has been comparatively neglected, despite the fact that incomplete data is common in real-world agricultural datasets. This makes it an important field of study with interesting applications in agriculture.

Most Influential Approaches That Inspired the Work

1. **Context Encoder by Pathak et al., 2016**: Established the use of adversarial training and autoencoders for image inpainting, enabling the prediction of missing areas in a picture based on the surrounding context. *Why interesting*: Similar to the goal of image outpainting, this work set the foundation for applying deep learning to problems involving the creation of missing or partial portions of an image.

2. **Lizuka et al. (2017)**: Presented Globally and Locally Consistent Image Completion. *Influence*: Suggested using both local and global discriminators to guarantee global coherence and local texture quality in generated images. *Why interesting*: Where both fine details and global structure need to align, this dual discriminator configuration is essential for producing realistic and cohesive outpainting results.

3. **Qingguo Xiao et al. (2020)**: Dense Residual Learning for Image Outpainting. *Impact*: Dense residual blocks and semi-U-Net architecture were added to improve the extraction of high-level features and preserve minute details in the outpainted regions. *Why interesting*: Dense residual learning offers an effective method of managing outpainting complexity by preserving network depth while retaining detailed information.

Literature Survey: State-of-the-Art (SOTA) Work

1. **Pathak et al.'s Context Encoder (2016)**: Context Encoders is one of the foundational works of image inpainting, introduced by Pathak et al. They used adversarial training in conjunction with autoencoders to forecast missing image segments based on the surrounding context. This work paved the way for related tasks like image outpainting and showed how deep learning can produce realistic image completions.

2. **Lizuka et al. (2017)**: Presented Globally and Locally Consistent Image Completion using a novel Generative Adversarial Network (GAN) with both local and global discriminators. While the local discriminator focuses on minute details, the global discriminator ensures the image is coherent overall. *Relevance*: This dual discriminator configuration is very important for outpainting, as peripheral images need to blend into the background and look realistic.

3. **Sabini and Rusak's Painting Outside the Box (2018)**: One of the first attempts to stretch images beyond their limits, focusing on image outpainting using GANs. Convolutional neural networks (CNNs) were trained adversarially to hallucinate picture material outside of the given borders.

4. **Qingguo Xiao et al. (2020)**: Dense Residual Learning for Image Outpainting. The method preserves specific information while improving high-level feature representation using semi-U-Net architecture and skip connections. This significantly increased the quality of outpainting, especially in terms of preserving sharpness and spatial consistency.

5. **Cheng et al.'s InOut: Diverse Image Outpainting using GAN Inversion (2022)**: Proposed GAN inversion to solve the diversity problem in image outpainting. The model produces distinct extrapolations by projecting the input image back into the latent space of the GAN. *Relevance*: This emphasis on diversity and many outputs is a fresh approach, offering methods for producing believable extensions, helpful for applications where environmental fluctuations must be considered.

Key Insights Driving Progress

Several key discoveries have driven advancements in deep learning-based image generation:

1. **Two Different Discriminators (Local and Global)**: Ensures that generated images are coherent at both fine detail (local) and overall structure (global) levels through the use of local and global discriminators (Iizuka et al., 2017). This addresses the challenge of creating aesthetically pleasing textures that also preserve the overall structure of the image.

2. **Dense Residual Learning**: Dense residual blocks, introduced by Xiao et al. (2020), addressed the vanishing gradient problem during training of deeper networks. Residual learning allowed the network to skip connections, preserving delicate information between layers and improving feature extraction for outpainting

tasks. *Impact*: This boosted model performance and increased the ability to produce realistic, high-quality outpainted regions.

3. **Stability in Training and Progressive Learning**: Cheng et al. (2022) proposed progressively increasing the task complexity throughout training (e.g., starting with smaller masks or simpler regions). *Impact*: This stabilises GAN training and ensures better performance when producing larger, more complex outpainted regions.

4. **GAN Inversion and Latent Space Exploration**: Models can create diverse realistic extensions by mapping the image back into the GAN's latent space (Cheng et al., 2022). This is useful for addressing the issue of deterministic outputs, ensuring variability in generated outpainted regions.

5. **Skip Connections and Semi-U-Net Architectures**: By using skip connections in U-Net and semi-U-Net architectures, models can retain both high- and low-level features during outpainting. *Impact*: This leads to crisper, more lifelike outputs, especially when handling complex textures.

6. **Attention Mechanisms and Feature Propagation**: Contextual attention techniques direct the model to relevant regions of the image, improving the generation of missing or extended components. *Impact*: These attention methods enhance the model's ability to produce coherent and semantically meaningful outpainted regions.

- Additionally, the use of dilated convolutions and style transfer techniques has shown potential for increasing neuron receptive fields and improving realism in image generation tasks.

- All these insights contribute to advancing work in the field of image outpainting.

Approach

Methodology

1. We will first downsample images for faster training.

2. Then we will split the dataset into Training and Validation sets.

3. Our preprocessing steps are as follows:

    - Normalize the training image.
    - Mask out the center portion of the image.
    - Compute mean pixel intensity over the unmasked region.
    - Set the outer pixel of each channel to the average value mean.
    - Concatenate the masked and normalized image.

4. **Architecture Description**:

    - Our method is based on performing image outpainting on crop datasets using Generative Adversarial Networks (GANs), specifically a DCGAN architecture. There are two primary components:

  - **Generator**: An encoder-decoder convolutional neural network (CNN) that predicts the image's expansion beyond predetermined borders.
  - **Discriminator**: A CNN that distinguishes between artificially produced outpainted images and actual cropped ones. The discriminator ensures both local texture quality and overall coherence using local and global components.

- The discriminator gives feedback, encouraging the generator to enhance the output's quality and realism, while the generator is trained to extrapolate the missing crop regions.
- **Three-phase training** is employed to stabilize the GAN training process:
  (a) **Phase 1**: Pre-training the generator using MSE loss to generate basic outpainted images.
  (b) **Phase 2**: Conditioning the discriminator to differentiate between real and generated images.
  (c) **Phase 3**: Adversarial training, where both the generator and discriminator are updated to improve outpainting quality.

Rationale for Advancing the State-of-the-Art

1. **Resolving Detail Loss and Spatial Coherence**:
   - **Limitation of Current Models**: Although skip connections and dual discriminators help maintain both local and global consistency, models often fail to capture minute features, particularly in areas distant from the source. This leads to hazy or unnatural textures.
   - **Our Approach**: To improve depth and preserve finer details during generation, we use dense residual learning to retain more high-level information. This ensures better texture generation and smoother transitions between the original and outpainted areas.
2. **Enhancing Stability of Training through Progressive Learning**:
   - **Current Drawback**: GAN training is frequently unstable, particularly when producing large outpainted regions. Models can collapse or generate low-quality images when faced with large extrapolation tasks.
   - **Our Approach**: We ensure stable GAN training and enhance the model's ability to handle larger, more complex extensions by utilizing progressive step learning, where the complexity of the task is gradually increased during training.
3. **Using GAN Inversion to Incorporate Diversity**:
   - **Limitation of Current Models**: Most outpainting techniques generate a single, predictable result, limiting their applicability to creative or ambiguous tasks requiring multiple plausible extensions.
   - **Our Approach**: By incorporating GAN inversion, we generate multiple outputs from a single input image by exploring different points in the latent space. This ensures high-quality, diverse extrapolations.

4. **Improving Feature Transfer with Semi-U-Net Structure**:
   - **Current Limitation**: Standard U-Net architectures with skip connections improve feature retention but introduce redundancy and high computational costs when applied to deep networks for outpainting tasks.
   - **Our Approach**: We employ a semi-U-Net architecture, where skip connections between the encoder and decoder's early layers are arranged more deliberately. This results in more accurate and efficient outpainted regions by maintaining important low- and high-frequency information without overloading the model.
5. **Customized Improvements for Agricultural Data Based on Application**:
   - **Present Drawback**: Most outpainting techniques are optimized for general-purpose datasets, like urban landscapes, and do not perform well on domain-specific data such as crop or agricultural photos, where fine color and texture details are crucial.
   - **Our Solution**: We tailor our technique specifically for crop data, ensuring that the model preserves important details such as leaf patterns, grain structure, and plant health indicators. This makes the approach highly useful for agricultural surveillance tasks.

Include Diagrams to Explain Your Reasoning or Approach

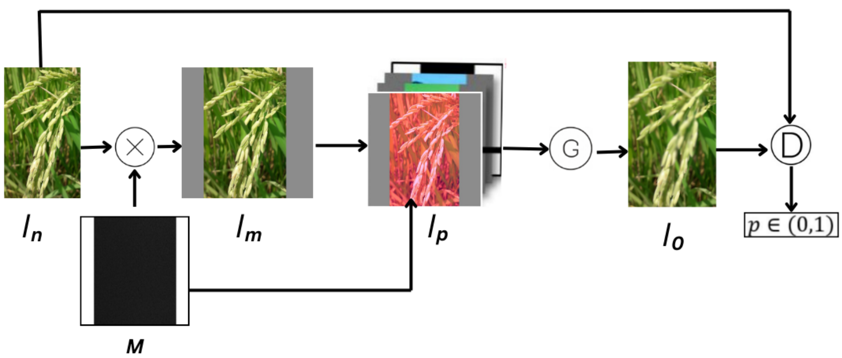

Figure 2: Training pipeline

We use the following notation to explain the image outpainting process:

- **In:** Input Image
- **M:** Mask
- **Im:** Masked Image
- **Io:** Output Image

This gives an overview of how an image is processed. First, the image is masked and passed through the generator, which produces the output image (Io). The discriminator then evaluates the generated image as either real or fake based on its resemblance to the original image.

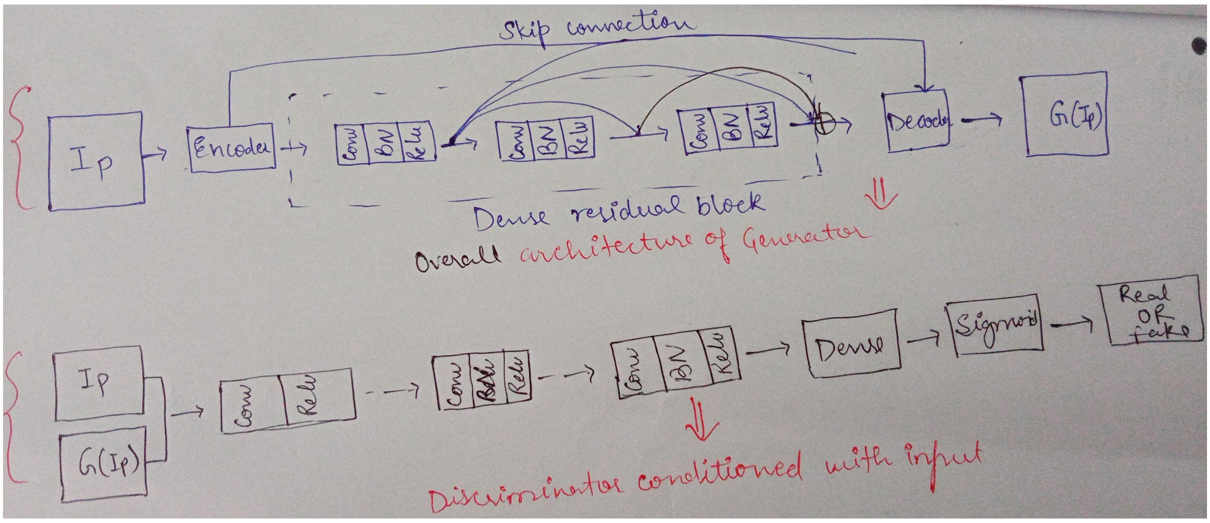

Figure 3: Training pipeline

The above given image provides a clear view of this architecture. The generator G follows the symmetric encoder-decoder baseline, and it is specifically a dense residual network. The encoder stage learns image features through a process of down-sampling and characterizing images. The decoder stage decodes the previously learned features and restores the image by upscaling the characterized maps.

Multiple Approaches and Justification

If you plan to try multiple approaches, describe your intuition for each one, and write them in the order of priority. We intend to test out the following strategies:

- **Local and Global Discriminators in DCGAN:** Our main strategy is to balance global coherence and local detail production using a DCGAN equipped with local and global discriminators. This method is based on Lizuka et al.'s work, where boundaries between the produced and original picture portions are better defined by local discriminators.

- **Generator Dilated Convolutions:** We shall use dilated convolutions to enhance the generator's receptive field. For applications like outpainting, where the model must hallucinate larger parts of the image, this method enables the model to "see" a larger portion of the visual context when creating new information.

- **Perceptual and Style Loss:** We will experiment with adding perceptual loss and style loss, inspired by recent developments in image inpainting. These loss functions encourage the model to produce textures and details that are more akin to actual crop photos, helping to improve the qualitative appearance of the generated images.

- **Hybrid Architecture:** We also want to try hybrid architectures such as DCGAN with Vision Transformer or DCGAN with semi-U-Net.

Justify why you think your methodology will work:

Our approach is built on proven techniques for creating realistic graphics using GAN-based image creation and image inpainting. Here's why we think our strategy will succeed:

- **Proven Effectiveness of GANs:** Tasks requiring the creation of visually realistic images are ideally suited for GANs, and the use of a discriminator and generator together guarantees that the quality of the images produced improves over time.
- **Local and Global Discriminators:** The model will be able to concentrate on distinct levels of image realism with the aid of local and global discriminators. Local discriminators will maintain texture quality at the boundaries, while global discriminators will ensure overall coherence.
- **Dilated Convolutions for Contextual Learning:** To generate believable out-painted regions, the generator needs to extract additional context from the input crop image. Dilated convolutions expand the receptive field without adding more parameters, helping the model "see" more of the image context.
- **Stabilized Adversarial Training:** The adversarial training process, which is often problematic in GAN-based models, is stabilized via three-phase training. To reduce the instability frequently linked to GAN training, we divide the generator pre-training and discriminator pre-training stages.

References

1. Y. Kim, J. Lim, and C. Kim, "Painting Outside As Inside: Edge Guided Image Outpainting via Bidirectional Rearrangement with Progressive Step Learning," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 993-1002.
2. C. Cheng, C. Lu, and X. Ren, "InOut: Diverse Image Outpainting via GAN Inversion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 14090-14100.
3. K. Wang, J. H. Pan, C. C. Loy, and L. Lin, "Diverse Image Completion with Bidirectional GANs," *IEEE Transactions on Image Processing*, vol. 29, pp. 3497-3506, 2020.
4. M. Mirzaei and M. Abbasi, "A hybrid machine learning approach for intelligent IoT energy management systems in smart buildings," *Neural Networks*, vol. 156, pp. 225-233, 2023.
5. S. Zor, B. Baykal, and O. Koncagul, "Deep neural networks in decision support systems: Application of multi-objective optimization in power allocation problem," *Expert Systems with Applications*, vol. 213, 2022.
6. A. Liu and Y. Huang, "Generative Adversarial Networks for Image Completion," Stanford University, CS230 Project Report, Spring 2018.