**Project Title:** Analysis of Effective Methods for Watermark Detection and Removal

**Team members:**

| Full name | ASU ID |
| --- | --- |
| Meesum Ali Khan | 1225453632 |
| Abichal Ghosh | 1225427294 |
| Nikhil Karthik Bindem | 1222520364 |
| Zuy Pham | 1216711899 |

**Problem statement:**

*Describe the problem you have chosen to investigate. (2-5 sentences)*

Watermarks are a standard way of protecting a person’s right to the photos and images they have created. However, with advancements in Deep Learning and other Machine Learning methods, it is now becoming possible to attack these watermarks, by detecting them and even removing them almost seamlessly from images. Our goal is to explore the effectiveness of Deep Learning methods against watermarks, by trying to do just that. Moreover, we will also work with different styles of watermarks, and try to find the ones that are least susceptible to detection and removal.

**Related work:**

*List at least 2 references to previous studies that are most relevant to your chosen problem (e.g., proposed solutions to address the same or similar problem). References should be written using the* [*ACM citation style*](https://www.acm.org/publications/authors/reference-formatting)*.*

1. Xinyun Chen, Wenxiao Wang, Chris Bender, Yiming Ding, Ruoxi Jia, Bo Li, and Dawn Song. 2021. REFIT: A Unified Watermark Removal Framework For Deep Learning Systems With Limited Data. In Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security (ASIA CCS '21). Association for Computing Machinery, New York, NY, USA, 321–335. <https://doi.org/10.1145/3433210.345307>
2. Shangwei Guo, Tianwei Zhang, Han Qiu, Yi Zeng, Tao Xiang, and Yang Liu. 2020. Fine-tuning Is Not Enough: A Simple yet Effective Watermark Removal Attack for DNN Models. <https://doi.org/10.48550/ARXIV.2009.08697>
3. Xiang Li, Chan Lu, Danni Cheng, Wei-Hong Li, Mei Cao, Bo Liu, Jiechao Ma, and Weishi Zheng. 2019. Towards Photo-Realistic Visible Watermark Removal with Conditional Generative Adversarial Networks. In International Conference on Image and Graphics.
4. Jing Liang, Li Niu, Fengjun Guo, Teng Long, and Liqing Zhang. 2021. Visible Watermark Removal via Self-calibrated Localization and Background Refinement. CoRR abs/2108.03581 (2021). arXiv:2108.03581 <https://arxiv.org/abs/2108.03581>

**Initial hypothesis:**

*What is the main research question you want to investigate, and what do you think your results will show? (1-2 sentences)*

We want to investigate different methods of detecting and removing watermarks from scenery images, and then compare which watermark types are most vulnerable to being detected and removed. We think that we will be able to successfully detect watermarks in most of the images, and watermarks that are less transparent, print-style fonts will be most vulnerable, while cursive style fonts and signatures will pose more of a challenge. We believe that Deep Learning will work best for filling in the pixels after removing the watermark.

**Dataset(s):**

*Describe which dataset(s) you will use for your analysis. How will you split the data into subsets for your experiments? What preprocessing or data cleaning will be required? How large is the dataset? What is the class distribution (if applicable)? Where will you obtain the dataset(s)? Use the table below to summarize this information. If you are proposing to create a new dataset, summarize your plan below and use the table to summarize the planned dataset.*

| Dataset source (link and reference) | Vesnin Dmitry. 2023. Scenery Watermark Detection. (January 2023). Retrieved February 26, 2023 from https://www.kaggle.com/datasets/qwertyforce/scenery-watermarks |
| --- | --- |
| Number of instances | 22762 images |
| Number of features | 512 x 512 x 3 = 786432  (Each image is 512 x 512 pixels and RGB-colored) |
| Class distribution (# instances in each class, if applicable) | Watermark: 9983 images (43.9%)  No watermark: 12779 images (56.1%) |
| Dataset splits | 60% Train, 20% Validation, 20% Test |
| Preprocessing steps | Basic image processing |

**Method(s):**

*What algorithm(s)/method(s) will you use to address the problem and why? What is novel about*

*your choice compared to what has been shown in previous work? What libraries or tools will you*

*use to implement your method(s)? How does your chosen method compare to the state of the*

*art for your problem? (3-5 sentences)*

Our given problem can be split into two major steps. First, we need to detect the ROI of a watermark in a given image. Once we are able to accurately locate the watermark, we need to fill in the watermark pixels with values that will blend in with the surrounding pixels. For the first step, we plan on using a base CNN Deep Learning model(autoencoder-like structure) for the detection of watermarks, as CNNs are the standard for visual data. We will experiment with different layers and kernels to see what works best. [3] uses a model which builds upon UNet, while [4] uses an autoencoder structure. We will also try to replicate the model of [3] and also try implementing an advanced version of UNet, U2Net which is much deeper in comparison to UNet. We will also experiment by adding skip connections in both the UNet and U2Net architecture and also experiment with atrous convolutions to better detect the watermarks.

For the second step, we will compare a variety of methods for Pixel Value Prediction, ranging from simple sequential averaging of surrounding values to using Deep Learning for finding edges and filling pixels appropriately. We will also use some newer methods we’ve read about on this topic, and see if they work well in this scenario. As for the tools and libraries, we plan on using OpenCV for the Image Processing part, and for the Deep Learning models, we will use Tensorflow or Pytorch. We may also use some pretrained models for pixel generation.

**Evaluation:**

*How will you quantitatively measure the performance of your solution? Will you do any qualitative (non-numeric) assessment? How will you compare your proposed method(s) to existing methods from prior work or other baseline approaches? Try to think beyond classic evaluation metrics (accuracy, AUC ROC, recall, precision, F1 score)—are there any evaluation criteria that would be particularly useful for evaluating your solution with respect to the chosen problem? (3-5 sentences)*

Given that the problem is split into 3 different steps, the following methods are used for evaluating the measure of performance of the solution on each of these steps-

1. Watermark Detection - The task of detection of watermarks is a segmentation task and can be thought of as the classification of pixels as watermarks and not watermarks. Hence, the detected watermarks will be compared with their masks using Binary Cross Entropy Loss. Here the images will be divided into the train, val, and test sets where accuracy, recall and precision will be important evaluation metrics.
2. Watermark Removal - To measure how accurately the model predicts pixels to create the watermark-removed version, we compare the reconstructed image with the base no-watermark image using methods such as Mean Squared Error(MSE), Root Mean Squared Error(RMSE) and structural similarity index measure (SSIM). Previous works of [4] and [3] have used PSNR, RMSE, DSSIM and SSIM as an index for measuring accuracy. [3] achieved a PSNR value of 30.86, while the model used by [4] had a PSNR value of 43.48, SSIM of 0.9959 and an RMSE of 2.15. The goal here is to train the model to get scores similar to the ones achieved in these papers.
3. Comparison - In the end, the goal of the project is to find the watermark which works the best(i.e. which DL methods are not able to reconstruct properly). The set of reconstructed images which have the lowest similarity would correspond to the best watermarks to use.

**Management plan:**

*Describe how you plan to manage the project implementation. How will you divide the work for this project between group members? How will group members hold each other accountable? How will group members communicate and collaborate throughout the project? (3-4 sentences)*

To manage the implementation of the project and make sure everyone is adding meaningful contributions to it, we will follow the suggestions given in the project guidelines document. For the first step of detecting the ROI of a watermark in an image, we will assign two group members to experiment with different layers and kernels using a CNN Deep Learning model, and the other two can work on image preprocessing and annotated data. For the second step of filling in the watermark pixels with surrounding values, we will have all 4 members work together on different methods, and see what gives the best results. We will also research and experiment with newer methods for this task. We will hold each other accountable by setting weekly goals and monitoring progress during our weekly meetings. We have also created a Slack group for communication and collaboration throughout the project, and will maintain a Github repository for version control. Although we initially planned to divide the second step among group members, we realized that it may be challenging, so we decided to have all members work together on this task to ensure the best outcome.