# Image Visible Masking/Watermark Removal with Graph Neural Network

## CS224W Winter 2023

Haishan Gao (hsgao@stanford.edu), Shaoyang Luo (sluo3@stanford.edu)

February 8, 2023

## 1 OBJECTIVE

Visible watermarks and image masking are commonly used on digital images for protecting copyright and deterring unauthorized use. Previous work has shown that they can be effectively removed using image processing or machine learning techniques with promising results. In this project, we propose the use of graph neural networks (GNNs) to address this problem. By modeling the relationship between pixels in an image as a graph, GNNs can be used to remove watermarks and masks while preserving the underlying image content.

Previous research on this task has primarily focused on two main approaches: image processing techniques and machine learning methods. Image processing techniques aim to alter the appearance of the watermarked image to make the watermark less noticeable (Pei & Zeng (2006)). However, these techniques often result in the degradation of the underlying image content.

Machine learning methods, on the other hand, have shown promising results in removing visible watermarks through generating an image without the watermark from a watermarked image, including the use of Convolutional neural networks (CNNs) (Cao et al. (2019)), and Generative adversarial networks (GANs) (Li et al. (2019)).

In recent years, graph neural networks (GNNs) have been introduced as a new approach to solve various image processing tasks. We aim to explore the performance of GNN models on image watermark masking removal in our project. Compared to deep CNNs and GANs with large parameter numbers and high model complexity, GNN models can exploit the underlying graphical strucutre of the images using a relatively smaller set of parameters and faster training iterations. To narrow down the scope of this project, we plan to use gray-scaled images of US license plates as our dataset.

## 2 DATASET

We will be using the US License Plates dataset from Kaggle (Dinçer (2021)) that includes 4462 clear images of license plates from all 50 states. All the images would be converted to grayscale to start with.


Figure 1: A picture of license plate (original v.s. grayscaled v.s. distorted)

**Input/Output Behavior**

Various distortions will be done to randomly selected images to form the test inputs. Output pictures will be compared against non-distorted grayscale images pixel by pixel to access restoration accuracy.

Since visually recognizable images do not require 100% accuracy, we will not chase perfect restoration, but pragmatic goals. Besides, not all distorted images are recoverable. And thus we will explore the application limitations such as what types of distortion is recoverable, and what types images are recoverable.

## 3 MODEL

For our task, we plan to employ Graph Neural Networks (GNNs), specifically Graph Convolutional Networks (GCNs) introduced by Bruna et al. (2013), due to their capability in effectively capturing the graphical structure present in our image data and addressing non-local dependencies between pixels. GCNs are particularly suitable for our task given that alphanumeric characters often exhibit structured shapes, such as circles and lines. To that end, we aim to extract these features through GCNs.

In our GNN formulation, we model the pixels as nodes and the connection to its neighbor pixels as the edges. We will experiment with different definitions of neighboring in our project. The objective of the model is to predict the node value, i.e. a float number representing the gray scale and minimize its difference with the actual node value.

As a starting point, we will implement a foundational GCN model with one hop neighbors and two CNN layers. The layer-wise propagation equation will be:

$$H^{(l+1)} = \mathrm{ReLU}(\hat{A}H^{(l)}W^{(l)})$$

$\hat{A} = D^{(-1/2)}AD^{(-1/2)}$ where $A$ is the binary symmetric adjacency matrix.

We will then explore other neighboring strategies and network architectures. Then we might further investigate the GCN in conjunction with the Label Propagation Algorithm (LPA) introduced by Wang & Leskovec (2020) if time permits.

## REFERENCES


Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

Zhiyi Cao, Shaozhang Niu, Jiwei Zhang, and Xinyi Wang. Generative adversarial networks model for visible watermark removal. *IET Image Processing*, 13(10):1783–1789, 2019.

Tolga Dinçer. Us license plates, Nov 2021. URL `https://www.kaggle.com/datasets/tolgadincer/us-license-plates`.

Xiang Li, Chan Lu, Danni Cheng, Wei-Hong Li, Mei Cao, Bo Liu, Jiechao Ma, and Wei-Shi Zheng. Towards photo-realistic visible watermark removal with conditional generative adversarial networks. In *Image and Graphics: 10th International Conference, ICIG 2019, Beijing, China, August 23–25, 2019, Proceedings, Part I 10*, pp. 345–356. Springer, 2019.

Soo-Chang Pei and Yi-Chong Zeng. A novel image recovery algorithm for visible watermarked images. *IEEE Transactions on information forensics and security*, 1(4):543–550, 2006.

Hongwei Wang and Jure Leskovec. Unifying graph convolutional neural networks and label propagation, 2020. URL `https://arxiv.org/abs/2002.06755`.