The University of Azad Jammu and Kashmir, Muzaffarabad

Department of Software Engineering

Computer Vision

Course Instructor: Engr Ahmed Khawaja

Roll no: 2022-SE-10-34-38

Semester: 7th

Session: 2022-26

CV Project Report

Project Report

PhysioNet - Digitization of ECG Images

Title:

A deep learning pipeline for digitizing ECG (electrocardiogram) images back into time-series signals using segmentation models.

Project Overview

The goal of this competition is to extract digital ECG signals from scanned ECG images (paper recordings). This solution uses a multi-stage deep learning pipeline:

- **Stage 0**: Key point detection and image normalization
- **Stage 1:** Grid detection and perspective rectification
- **Stage 2:** Pixel-level waveform segmentation and signal extraction

Abstract:

This research presents a deep learning–based approach for the digitization of electrocardiogram (ECG) images into machine-readable signals. Using ECG image data, the system preprocesses and normalizes images, corrects perspective and grid alignment, and applies convolutional neural network models to extract waveform traces. The methodology includes image normalization, rectification, signal extraction, and ensemble-based prediction to reconstruct multi-lead ECG time series. The results demonstrate that automated digitization can accurately convert scanned ECG images into usable digital signals, enabling their integration into clinical analysis and machine learning workflows.

Keywords:

ECG digitization, electrocardiogram images, deep learning, image preprocessing, signal extraction, convolutional neural networks, time-series reconstruction, medical image analysis, healthcare AI, PhysioNet

1. Introduction:

Electrocardiograms (ECGs) are essential for diagnosing heart conditions, yet many records remain in paper or image form, limiting their use in digital analysis. Manual digitization is slow and error-prone, creating a need for automated solutions. Recent advances in deep learning enable accurate extraction of waveform information from images. This study proposes a deep learning-based pipeline that preprocesses and normalizes ECG images, corrects perspective and grid alignment, and extracts waveform signals to reconstruct multi-lead digital ECG time series, supporting efficient clinical analysis and machine learning applications.

2. Dataset Description:

The dataset for this study is taken from the *PhysioNet ECG Image Digitization (2025)* Kaggle competition. It consists of ECG images and metadata used to train and evaluate models for converting ECG images into digital signals. The dataset includes the following files:

- **train.csv** – contains training information and labels for ECG samples
- **test.csv** – contains test ECG image IDs for which predictions must be generated
- **train/** – folder containing training ECG images
- **test/** – folder containing test ECG images
- **sample_submission. parquet** – a sample file showing the correct submission format

The objective is to build a model that digitizes ECG images and generates accurate predictions for the test set in the required submission format.

3. Preprocessing

Two approaches were used to prepare ECG images. First, rotated images were transformed to 3200×2400 resolution using a homography matrix and grid remapping. Second, grid points were rescaled via the homography matrix and remapped directly on rotated images. In both cases, the lower ECG-signal region was cropped and the top personal-information region removed. Images were then converted to grayscale and combined with coordinate-based features to provide both visual and spatial information for accurate prediction.

4. Methodology

4.1. Training Strategy:

The models were trained using rectified ECG images as input. The training configuration included:

- **Input Parameters:** Stage-1 rectified images; whole model input shape (3, 1280, 5600) and series model input shape (4, 3, 480, 5600).
- **Loss Function:** BCEWithLogitsLoss with positive class weight set to 20 to handle class imbalance.
- **Training Setup:** 50 epochs with batch size 4.
- **Optimization:** AdamW optimizer with learning rate ranging from 5e-4 to 1e-3 and weight decay of 0.01.

- **Scheduler:** CosineAnnealingLR used to adjust the learning rate during training.

4.2. Model Training:

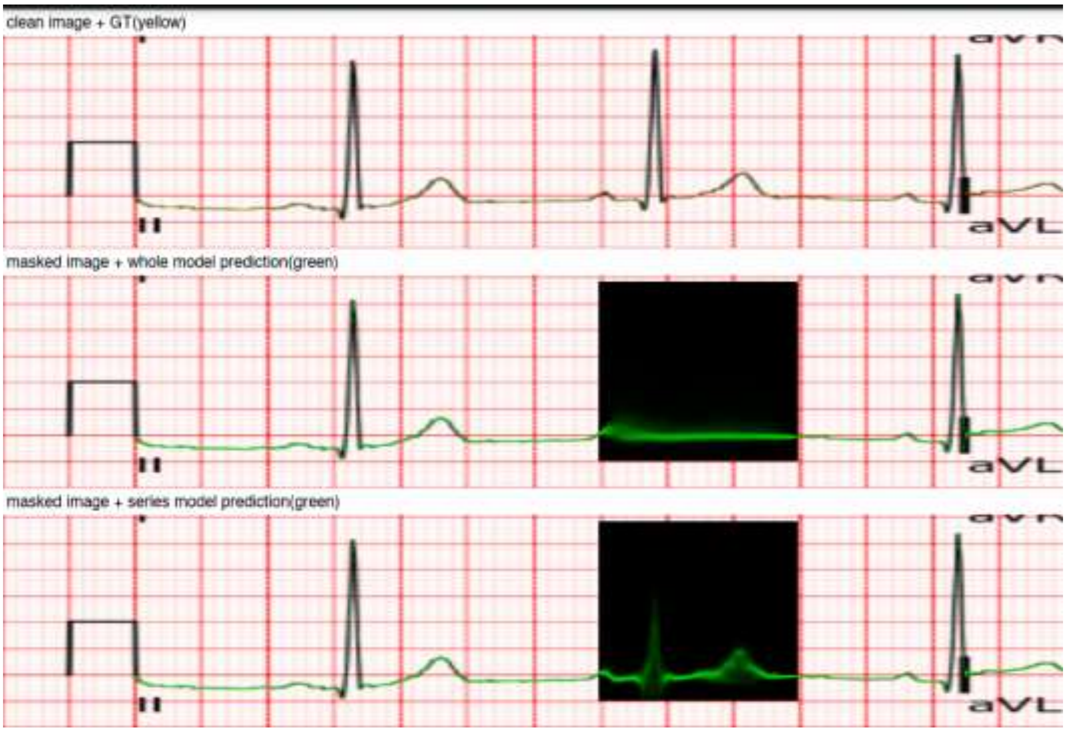
The model training follows a multi-stage deep learning framework where separate neural networks are trained for ECG image normalization, rectification, and waveform extraction. The steps included:

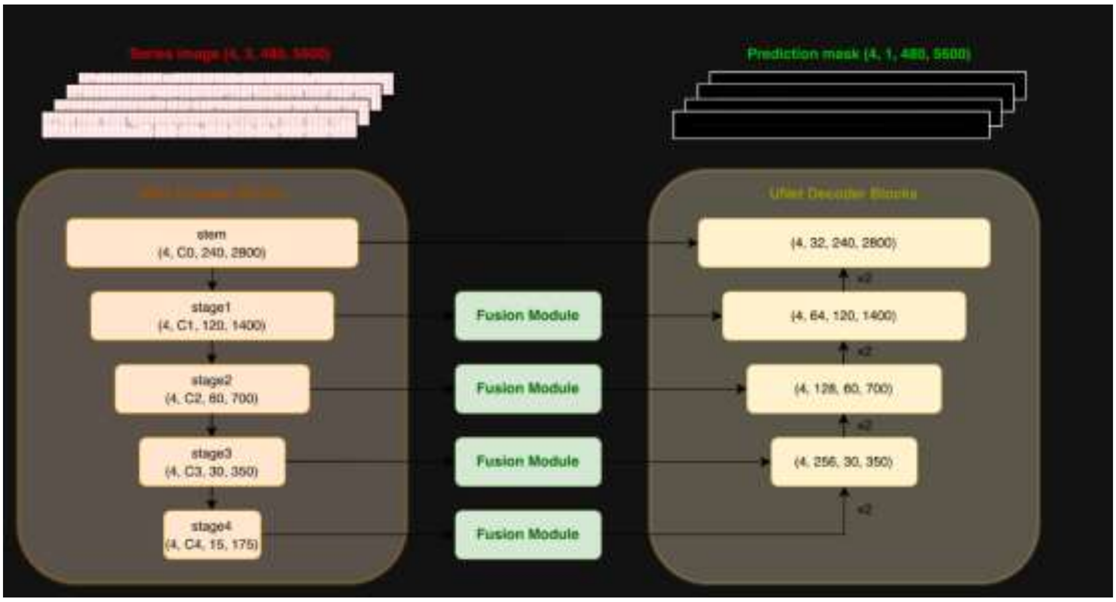
- Preparing the training data by loading ECG images, resizing them to a fixed resolution, and generating ground-truth annotations such as keypoints, grid locations, and waveform masks.
- Training Stage-0 and Stage-1 models to learn image normalization and grid/lead alignment using convolutional neural networks and supervised learning.
- Training Stage-2 segmentation models (whole-image and lead-specific networks) to predict pixel-level ECG traces, using data augmentation and validation to improve generalization.
- Saving trained weights and checkpoints for each stage, which are later loaded during inference to generate final predictions.

5. Models:

Two models were implemented for ECG segmentation:

- **Whole model:** A U-Net architecture with timm encoder and custom decoder. ECG images were cropped and resized, then passed through the network to generate segmentation masks.
- **Series model:** A 2.5D U-Net that takes four ECG series as input to capture inter-lead relationships. Each series was cropped around the zero-baseline and processed using a shared encoder, with features fused in decoder connections.
- **Fusion module:** Conv2D, shared Conv2D, and Conv3D fusion methods were tested to combine series features. Conv2D fusion achieved the best validation performance and was used in the final model.





6. Rectification

Two primary enhancements were implemented for the original rectification method to improve output quality

- Grid point positions are calculated using a weighted average from the feature map to preserve floating-point precision.
- Surface fitting is utilized to identify and remove outliers, with the resulting gaps filled through interpolation or the regression function itself. This approach ensures smooth image edges and significantly enhances the overall structural integrity of the rectified images.

7. Results

An ensemble-based inference pipeline was used. Multiple pretrained whole-image and series models were combined with test-time augmentation (TTA) and averaging to improve waveform prediction accuracy.

Model	Backbone	Type	Input	Output	Prediction Strategy	Ensemble
1	EfficientNet-B7	Whole model	Rectified ECG image	Segmentation mask	Whole image inference	✓
2	EfficientNetV2-L	Whole model	Rectified ECG image	Segmentation mask	Whole image inference	✓
3	EfficientNet-B6	Series model	Lead images	Signal mask	Shared conv2d fusion	✓
4	EfficientNet-B6	Series model	Lead images	Signal mask	Shared conv2d fusion	✓
5	EfficientNetV2-L	Series model	Lead images	Signal mask	Conv3D fusion	✓
6	EfficientNetV2-L	Series model	Lead images	Signal mask	Conv2D fusion	✓

7. Conclusion

This study demonstrates the effectiveness of deep learning–based models for ECG image digitization and waveform reconstruction. The proposed multi-stage pipeline with preprocessing, segmentation, and ensemble prediction successfully converted ECG images into accurate digital signals. Future work can focus on incorporating more advanced architectures, improving signal post-processing, and optimizing ensemble strategies to further enhance prediction accuracy and robustness.