Experience in AI

Khan Fashee Monowar
E-mail: khanfmonowar@gmail.com
Website:
khanfasheemonowar.github.io

November 2023

I have summarized my noteworthy academic, research, and industrial projects here.

1 Industry Experience

- **Intelligent FIFAe Livestream Object-aware Cropping System**
  - **Project Timeline:** April 2023 - October 2023
  - **Technology:** Python, Computer Vision, Signal Processing.
  - **Role:** Research Data Scientist
  - **My Responsibilities:**
    - Led the full projects.
    - Conducted Research to find the solution.
    - Documented the findings and suggested a new methodology.
    - Collected FIFAe livestream videos and annotated initial Dataset.
    - Trained cutting-edge object detection model
    - Fine-tuned hyperparameters.
    - Developed a **novel algorithm** for intelligent and object-aware video cropping which takes model predictions as input.
    - Implemented signal processing techniques to increase user experience
    - Developed end-to-end solution.
    - Preparing Test cases & Performing Testing.
    - Documenting Technical details.

  In this project, our objective was to create a system that automatically crops live-streaming FIFAe videos, ensuring that the soccer ball remains consistently visible in the frames. The goal was to provide an excellent user experience, allowing users to enjoy relevant and engaging scenes(celebration, attacking, kickoff, replay, etc) optimized for vertical **smartphone video (portrait) mode**, specifically for a social media platform. To achieve this, we have done extensive research into the detection of **small objects using object detection models** and also to develop cropping methods.

To realize our objectives, we initiated the project by **preparing vast image data from FIFAe videos and annotating 1000** images using **Roboflow**. Then, we trained a cutting-edge **object detection model** using the **transfer learning** method. After that, we fine-tuned the model's parameters to optimize its performance and prepared it for inference. Later, we developed **our unique cropping method** utilizing **signal processing techniques** which takes the object detection model inference result as input. **The cropping algorithm** provides a seamless and contextually relevant cropped video experience, with a continuous focus on the appropriate scenes. Later, through an iterative process involving trial and error, we refined our algorithm, progressively enhancing its capabilities. Once we achieved our desired results, we finalized the **end-to-end process** and delivered it for production.

- **Object Detection in FIFAe Game Livestream Using YOLOv7**
  - **Project Timeline:** March 2023 - April 2023
  - **Technology:** Python, PyTorch, Object Detection.
  - **Role:** Research Data Scientist
  - **My Responsibilities:**
    - Led the project.
    - Investigated prevailing systems.
    - Documented the findings and proposed a system.
    - Gathered FIFAe images and annotated the dataset.
    - Trained the YOLOv7 model with transfer learning
    - Testing and Evaluation
    - Production Deployment with Comprehensive Documentation

  In our project, the objective was to detect multiple objects (ball, players, referee, goalkeepers, goalpost, score layouts etc) in FIFAe game streaming. To accomplish this, we began by gathering images from FIFA video games and creating an annotated dataset with **1500** image samples with a total of **34,864 annotation**s. Subsequently, we employed **the transfer learning technique to train the YOLOv7 model** (cutting-edge object detection model) using our dataset, which consisted of **9 distinct classes**. Additionally, we **fine-tuned hyperparameters** and adjusted image sizes to enhance the model's accuracy. Following the training process, we assessed the model's performance in real video scenarios. Later, we sent the model to production team.

- **PaddleOCR Detection Model Custom Training**
  - **Project Timeline:** December 2022 - February 2023

- **Technology:** Python, Deep Learning, OCR.
- **Role:** Research Data Scientist
- **My Responsibilities:**
  * Investigated scenarios where PaddleOCR did not meet our requirements
  * Documented the findings.
  * Collected relevant data and annotated initial dataset.
  * Conducted training for the PaddleOCR Det model.
  * Fine-tuned hyperparameters.
  * Assessed the model's performance.
  * Documented technical specifics.
  * Transferred the model to the production team.

In this project, **we enhanced the accuracy of PaddleOCR** in text extraction for specific scenarios by training the **PaddleOCR detection model** on a **custom dataset**. We carefully **gathered and annotated** over 6000 data using the **PaddleOCR annotation tool**. Through this process, we closely monitored its performance and adjusted hyperparameters to optimize its capabilities. We followed the official PaddleOCR documentation and implemented the **knowledge distillation training** approach. After fine-tuning the model, we evaluated its performance against our test cases. Subsequently, **we integrated the trained model into our production environment**, ensuring thorough documentation for seamless deployment and ongoing usage.

- **Action Spotting in FIFAe Game:**
  - **Project Timeline:** September 2022 - November 2022
  - **Technology:** Python, Deep Learning, Computer Vision.
  - **Role:** Research Data Scientist
  - **My Responsibilities:**
    * Investigated existing research.
    * Selected a methodology based on the findings.
    * Implemented modifications to the selected methodology for optimization.
    * Collected relevant video data and annotated initial dataset.
    * Conducted training for video data.
    * Fine-tuned hyperparameters.
    * Assessed the model's performance.
    * Documented technical specifics.
    * Transferred the model to the production team.

We conducted thorough research to develop a system for **extracting time frames of game events** (such as kick-off, scoring, attacking, corner kicks, etc.) from FIFAe live stream videos. Following our research, we adopted the **Soccernet-v2 action spotting architecture**, which is tailored for the task of identifying and localizing particular actions or events within lengthy video sequences. This architecture seemed well-suited to our objectives. To gather the necessary data, we collected and annotated **200 FIFA esports game videos** sourced from platforms like YouTube and others. We employed the **Soccernet annotation tool** to streamline this process. Subsequently, we fine-tuned the model and optimized hyperparameters to achieve the highest level of accuracy possible. Following this, we evaluated the model's performance using a dedicated test dataset, and the results met our production standards, providing satisfactory accuracy levels for **version 1.0**. With comprehensive documentation in place, we successfully deployed the model into our production environment.

- **Bengali Handwritten OCR Development:**
  - **Project Timeline:** october 2021 - May 2022
  - **Technology:** Python, Deep Learning, Computer Vision.
  - **Role:** AI Programmer
  - **My Responsibilities:**
    - Investigated existing research on Bengali Handwritten OCR.
    - Choose Initial Strategy.
    - Collected 250,000 character images manually and clean the initial dataset.
    - Prepared dataset and conducted training with ResNet152V2.
    - Fine-tuned hyperparameters.
    - Assessed the model's performance.

  Our venture initiated a research and development project with the goal of pioneering the first dependable full-scale Bengali handwritten OCR system in Bangladesh. Our journey commenced with the challenging task of **classifying individual characters**, a challenge magnified by the complex nature of the Bengali letters and handwritten pattern, which encompasses vowels, consonants, and consonant conjuncts. To build our dataset, we gathered handwritten characters from our office's employees, amassing an extensive collection of over **250,000** character images categorized into **250 distinct classes**. During the data preparation phase, we applied a range of OpenCV methods and image filtering techniques to enhance the quality of our image dataset. Leading the charge, I played a central role in managing the data **collection, processing, and character classification** training phases. Additionally, we conducted a comprehensive review of existing research to choose our approach. Our strategy included the

utilization of a **ResNet152V2** model for training our multiclass dataset, yielding an impressive **81%** overall classification accuracy.

- **Information Extraction from Document Image:**
  - **Project Timeline:** March 2021 - November 2021
  - **Technology:** Python, Deep Learning, NLP.
  - **Role:** AI Programmer
  - **My Responsibilities:**
    * Gathered data using OCR.
    * Processed the text files for annotation.
    * Used Doccano tool for annotation.
    * Conducted Name Entity Recognition training using Spacy framework.
    * Fine-tuned hyperparameters.
    * Assessed the model's performance.

  In this project, our target was to extract entities from structured and unstructured document image data. We gathered a dataset and extracted text from structured and unstructured document images using English OCR, as well as **Amazon Textract**. Subsequently, we annotated the text data using the **Doccano** annotation tool to prepare for **name entity recognition (NER)** training. In this process, we had **7 classes** and worked with **300 text files** for training. We employed the **Spacy** framework to train the **NER transformer** model, fine-tuning the model's hyperparameters to achieve the best possible results. However, due to limited data availability, we reached a modest accuracy of **52%**. As this was a research project, we did not proceed with production implementation.

2 Project Experience

I have completed several projects and attended competitions for hands-on experience. Here are some of them worth mentioning-

- **NIH Chest X-ray-14 Multilabel Multiclass Classification**
  * **Technology:** Python, Computer Vision.
  * **Accomplishments:**
    · Built a Computer-Aided Diagnostic (CAD) model.
    · Worked with a dataset comprising 112,000 chest X-ray images.

- Trained MobileNet using a multilabel multiclass dataset with 14 classes.
- Conducted various hyperparameter tuning experiments to optimize model performance.
- Observed the impact of training and testing time augmentation on model performance.
- Applied a weighted binary cross-entropy loss function to reduce class imbalance.
- Calculated class wise accuracy, recall, precision, sensitivity, specificity, PPV, NPV and AUC scores.
- Achieved a mean accuracy of **84.95%** and a mean AUC score of **78.67%**.

In our research project, we developed a CNN-based diagnostic model with the goal of identifying **14 distinct diseases** from chest X-rays. For that, we have used NIH Chest X-ray-14 **Multilabel Multiclass** dataset. We conducted an exploration of various CNN architectures for this task, and among them, **MobileNet** proved to be the most effective, achieving an average classification accuracy of **84.95%**. Given our intention to deploy the trained model on edge devices, we opted for a lightweight architecture, such as MobileNet, to ensure efficient performance. To improve the model's performance, we employed both **training and test-time data augmentation** techniques. Using the MobileNet architecture, we processed a dataset that consisted of over 112,000 chest X-ray images, with the model being trained on **80,000** of these images. During the training process, we utilized a **weighted binary cross-entropy loss function** to optimize the model's performance and also took care of the class imbalance problem. [Link]

– **Pneumonia Detection using CNN with Explainable AI**

* **Technology:** Python, Computer Vision, Explainable AI.
* **Accomplishments:**
  - Built CAD model to detect pediatric pneumonia.
  - Trained MobileNet on the binary dataset
  - Conducted various hyperparameter tuning, and data augmentation.
  - Calculated class weights to deal with class imbalance
  - Apply **Grad-CAM** to generate heatmaps
  - Achieved a accuracy of **86.0%** and a recall score of **99.06%**.

In this research project, I developed a CAD model to identify pediatric pneumonia in chest X-ray images. To do this, I utilized a **binary chest X-ray dataset** from Kaggle, consisting of **normal**

and pneumonia classes. The dataset posed a class imbalance challenge, with a total of **4,642 training** images, including 3,418 pneumonia images and 1,224 normal class images. To address this issue, we implemented the **Sklearn class_weight method** to assign class weights, which were then used in the **fit_generator()** function from the Keras library. To improve the model's generalization ability, we implemented training-time **augmentation** techniques. Our choice of architecture was **MobileNet**, and we further enhanced the model's performance by fine-tuning the image size. The results were promising, with an achieved testing accuracy of **86%** and an impressive recall score of **99.06%**. In addition to model performance, we also focused on model interpretability. We leveraged the Explainable AI tool, **Grad-CAM**, to generate **heatmaps** that highlight the regions in the images that had the most significant influence on the model's decision. [Link]

– **AI-based Dhaka Traffic Detection Challenge in 2020**

  * **Technology:** Python, Computer Vision.
  * **Accomplishments:**
    - Attended an Object detention competition in Dhaka.
    - Collected and annotated images to manually expand the 21-class dataset.
    - Train yolov5 from scratch
    - Conducted various hyperparameter tuning
    - Achieved **53.4%** average mAP

  My team and I took part in the AI-based Dhaka Traffic Detection Challenge in 2020, which focused on evaluating the capabilities of cutting-edge methods for detecting and recognizing traffic vehicles on the roads of Bangladesh. The challenge encompassed **21 distinct traffic-related object classes**. To strengthen our dataset legally, we undertook the task of collecting and annotating additional images. Subsequently, we embarked on training the **YOLOv5** model, fine-tuning it with a range of hyperparameters and augmentations to optimize its performance. We achieved **53.4%** average mAP. [Link]

– **Twitter Data Sentiment Analysis**

  * **Technology:** Python, NLP.
  * **Accomplishments:**
    - Conducted comprehensive research on customer sentiment within the US airline industry.
    - Merged several datasets to create a robust dataset, training the algorithm on a total of **182,330** tweet data entries.

· Implemented LSTM model to enhance performance and accuracy in sentiment analysis. Introduced dropout layers to the models to reduce overfitting.
· Worked with the dataset featuring two classes: positive and negative sentiments.
· Achieved a notable 90.06% F1 score, indicating a high level of performance in sentiment analysis.

In this project, I conducted sentiment analysis using Twitter data related to US Airlines. The dataset was transformed into two classes: positive and negative sentiment. I developed a bidirectional LSTM model from the ground up, leveraging 182,330 tweets for training. To address overfitting, I implemented the dropout technique. Throughout this project, I gained hands-on experience in fundamental NLP concepts including Tokenization, Text Vectorization, TF-IDF, Word Embedding, and more. Notably, the model achieved around a 91% F1 Score.[Link]

3 Research Experience & Publications

In my final year of undergraduate studies, I collaborated with multiple professors. Along with Professor Dr. Al Mehedi Hasan and Dr. Jungpil Shin, I also worked with Professor Dr. Ali Hossain from the Department of Computer Science and Engineering at RUET. Here are the details of my publications in brief:

– **Lung opacity classification with convolutional neural networks using chest x-rays [ICECE 2020]** [Link]
Authors: **Khan Fashee Monowar**, Md Al Mehedi Hasan, and Jungpil Shin

Our Contributions:
* Compared off-the-shelf CNN architecture's performance on the binary and multiclass dataset.
* Implemented weighted binary cross-entropy loss function to reduce with class imbalance.
* Phase 1 evaluates how well CNN architectures differentiate the Lung Opacity class from No Lung Opacity + Normal.
* Phase 2 evaluates how well CNN architectures differentiate the Lung Opacity class from only the normal class.
* In both cases, Xception achieved the best AUC scores, with **91.0% in Phase 1 and 99.1% in Phase 2**, respectively.

Short Abstract: Recently, deep learning achieves radiologist-level performance in chest X-ray interpretation. In our research, we trained and evaluated various deep convolutional neural networks (CNN) architectures to detect potential lung opacity from chest X-rays. We observed how strongly architectures could differentiate lung opacity from normal and other abnormal chest X-rays. In these circumstances, A CNN based model (Xception) achieved 91.0% AUC along with 83.95% accuracy. Moreover, we also observed models achieved a better performance on lung opacity vs normal chest X-ray classification (excluding abnormal class) where Xception achieved 99.1% AUC, 97.19% sensitivity, and 95.71%accuracy. Therefore, the purpose of this study is to investigate the classification ability of deep CNN architectures which helps to develop an automatic lung opacity detection system.

– **A Lightweight Convolutional Neural Network Model for Child Pneumonia Classification [ICICT4SD 2020]** [Link]

Authors: **Khan Fashee Monowar**, Md Al Mehedi Hasan, and Jungpil Shin

Our Contributions:

* Designed a **lightweight, 6 layers** CNN architecture that has 6.8 million parameters
* Trained binary dataset (pneumonia and normal class) with the model from scratch
* Compared our model performance with another off-the-shelf CNN models and recent studies
* Achieved the best Accuracy, Precision, F1, Specificity and AUC score among the studies we have mentioned.

Short Abstract: Pneumonia is still a serious threat for children including newborns. In recent years, Computer-aided detection system (CAD) and medical image classification are progressively turning into another research territory. Currently, Researchers build various models to detect pneumonia from chest X-rays. However, there is still a lack of computationally efficient models to diagnose pediatric pneumonia. Further, some off-the-shelf or pre-trained models are not always suitable for mobile and embedded vision applications since these models are not lightweight. In our research, we built a lightweight CNN model from scratch using basic building blocks which able to learn lung texture features and detect pediatric pneumonia. We compared our proposed model performance with some off-the-shelf models. Our proposed model achieved the best AUC (99.0%), test accuracy (94.6%), F1 (94.7%), precision (93.2%) and

specificity (93.1%) scores. Moreover, We employed several data augmentation algorithms to increase the model's classification ability.

– **ECG Heartbeat Classification Using Ensemble of Efficient Machine Learning Approaches on Imbalanced Datasets [ICAICT 2020]** [Link]

Authors: Md Atik Ahamed, Kazi Amit Hasan, **Khan Fashee Monowar**, Nowfel Mashnoor, and Md Ali Hossain

Our Contributions:

* We compared some well-known machine learning approaches with other state-of-the-art related methodologies
* Used two datasets, MIT-BIH Arrhythmia which contains five classes and PTB Diagnostic ECG which contains two classes.
* We also applied an ensemble approach (hard voting)
* For MIT-BIH, ANN with class weight approach gained the best accuracy (**98.06%**).
* For PTB Diagnostic ECG , Ensemble method achieved the best accuracy, recall and precision (**97.66%, 96.90% and 97.06% respectively.**)

Short Abstract: Being electrocardiogram already an established method for analyzing cardiac health, it gained many researchers' interests to classify heartbeats accurately. In this paper, some well-known machine learning approaches are used by tuning and compared with other state-of-the-art related methodologies. The datasets are used in this research work, are highly imbalanced and handled with penalizing the loss value of the Artificial Neural Network (ANN) by assigning class weights. Two different enriched ECG datasets are selected for this research. They are MIT-BIH Arrhythmia which contains five classes and PTB Diagnostic ECG which contains two classes. About **98.06% and 97.664%** accuracy are achieved with proposed approaches for MIT-BIH Arrhythmia and PTB Diagnostic ECG dataset respectively. Both cases this research outperforms all the other state-of-the-art methodologies.