EAS 560 Internship Report

(Summer 2023)

By

Roshni Balasubramanian

50483779

roshniba@buffalo.edu

Table of Contents

1. Introduction

Third Estate Ventures is a firm based in Buffalo, focusing on noteworthily addressing issues prevalent in the community. The internship opportunity at the aforementioned firm offered the ability to function as a multi-project intern, who had to skillfully manage various Machine Learning (ML) and Computer Vision (CV) focused components across several projects undertaken by the firm. The said endeavor spanned roughly around three months starting from May 21st to August 18th, encompassing an average weekly commitment of 30 hours. This experience proved to be a resounding success, meticulously guided and overseen by the astute mentorship of Mr. Joseph Trapp and Ms. Lily Keane.

The assigned tasks were diligently undertaken in collaboration with Ms. Meghna Pal (50475251). While the core substance would remain consistent, the submitted reports would stand out as distinct renditions, thereby reflecting unique perspectives and articulation.

The first project of this commitment centered on Peoria County within the city of Peoria, situated in Illinois. This opportunity was endowed with the prospect of engaging in the realm of Natural Language Processing (NLP) and Optical Character Recognition (OCR). Notably a government initiative, the overarching objective encompassed the comprehensive overhaul of the county website dashboard. The prevailed challenge laid in transforming the dashboard characterized by scanty and haphazard information into an information-rich resource, akin to the Buffalo City website. The redesigned dashboard was meticulously crafted to present a diverse array of data, inclusive of but not restricted to data pertaining to commercial and official edifices, ownership records of buildings, real estate agents in the county, county officials, and non-profit organizations. The essence of this initiative resided in fostering enhanced engagement among citizens, particularly those with existing businesses or with a vested interest in establishing new enterprises in the county or within the periphery of the county.

Amidst this assignment, a subsequent transition marked the commencement of the second project, with a strong emphasis on CV and ML. The aim of the initiative is to locate and identify the anomalies on the exteriors of residential buildings that are deemed as violations by the law. Through the process of model training, an automated system was cultivated with the capability to facilitate engagement of the residents. This interactive platform empowered them with the ability to upload images depicting potential violations. The underlying mechanism of the system then rigorously assessed these images to ascertain the presence of violations and, subsequently, to determine their specific violation category. This automation negated the need for human intervention, consequently expediting and streamlining the process in a manner that is both efficient and less labor-intensive. By avoiding the necessity for manual human intervention, potential bottlenecks and administrative delays can also be minimized. The successful completion of the task was significantly facilitated by the utilization of Convolutional Neural Networks (CNNs), a pivotal technology renowned for its remarkable efficacy in this context.

2. Background

As previously elucidated, this experiential endeavor bestowed a comprehensive platform for engaging with an array of technical dimensions, ranging from rudimentary Python coding to the intricacies of advanced concepts such as NLP, OCR, CNN, etc.

The Peoria project encompassed a strategic orchestration of the *openpyxl* and *xml.etree.ElementTree* libraries, serving as foundational pillars for the purposeful structuring of Extensible Markup Language (XML) forms into Excel spreadsheets. Additionally, the integration of the *pytesseract* and Pillow Python libraries was pivotal, as these technologies effectively enabled the transformation of images into text through OCR methodology. The culminating touch was provided by the *nltk* package, which undertook the role of a proficient tokenizer, thereby facilitating NLP operations.

Conversely, the Violation Detection project navigated the realm of image-based data classification, a task necessitating meticulous data pre-processing. This involved a systematic sequence of preparatory measures.

Subsequently, the resultant dataset was funneled into *Rekognition*, a cloud-based ML model offered Amazon Web Services (AWS) established model, marking a preliminary classification step.

However, the pursuit of enhanced efficacy prompted use of different CNN model trials to supersede the legacy model. The initial exploration of fundamental model architectures encompassed the evaluation of two prominent frameworks: the *Visual Geometry Group-16* (*VGG-16*) and the *Residual Network (ResNet-50)*. This preliminary testing phase laid the foundation for subsequent model customization endeavors. The essence of this customization hinged on the integration of *Class Activation Mapping* (CAM) and the strategic incorporation of *Residual Blocks*; intricacies that were seamlessly achievable due to the flexibility afforded by the *PyTorch* framework. This synergy of the architectural components conferred an enhancement upon the resultant model of its classification capabilities.

Furthermore, an incisive comparative analysis is presented, offering insights into the contrasting performance of the legacy model versus the proposed model trials. Key parameters of evaluation include accuracy metrics, discernible drawbacks, and an exploration of avenues for future developmental pursuits.

The pedagogy of the Machine Learning course (CSE 574) offered the exposure to architectures such as *ResNet-50* and *VGG-16*. Assignments in pertinence to the same course provided the requisite knowledge foundation towards the mastery of customizing CNNs with *PyTorch*. The process of attending the informative sessions, integral to Programming and Database Fundamentals for Data Scientists course (EAS 503) contributed to the proficiency in Python programming, extending to file manipulation and other basic coding constructs.

3. Methods

3.1 Peoria County Project

3.1.1 Data Collection

The nucleus of the undertaken task revolved around the intricate and systematic collection of Form 990s, a pivotal and comprehensive document, from two distinct repositories, namely the Internal Revenue Service (IRS) website [1] and the Candid GuideStar website [2].

Within the domain of the IRS website, the modus operandi for access was facilitated through the strategic mechanism of bulk downloads. The crux of this approach lay in the acquisition of an expansive corpus of Form 990s that had been meticulously filed in the digital realm. This repository, similar to a bunch of nonprofit organization (NPO) financial disclosures, bore paramount significance due to its inherent timeliness and relevance. The focal point of this pursuit rested on capturing the very latest filings, encapsulated within the documentation pertinent to the year 2022. These filings, meticulously structured in the form of XML documents, represent a digital imprint of the recorded forms.

The process of procurement and consolidation culminated in the assembly of an extensive compendium of Form 990s. This vast reservoir of documents was systematically organized and securely ensconced within localized storage repositories, poised to serve as the bedrock for subsequent analytical explorations. The core essence of this strategy was to amass a comprehensive collection, granting a glimpse into the realm of NPO documents for the specified temporal span.

Conversely, the Candid GuideStar website emerged as yet another treasure trove of NPO-centric information, albeit veiled in a distinct format. Within this repository, the documents assumed the guise of scanned artifacts, delicately encapsulated within Portable Document Format (PDF) files. This repository, aptly designated as a veritable data hub for NPO information, encompassed a multifaceted array of data enclaves, among which were the coveted Form 990s. Noteworthy in this context was the characteristic feature of these documents scanned renderings that emanated an image-centric disposition within the PDF confines. The course of acquisition encompassed the delicate extraction of these scanned artifacts, ultimately infusing the project with a rich source of information, insight, and value derived from this invaluable resource. The dual nature of these repositories underscores the multifaceted nature of data collection, underscoring robust methodology of the project in embracing diverse information sources. By harmonizing these repositories, the project leveraged a versatile spectrum of data, seamlessly interweaving XML-structured Form 990s with scanned artifacts, effectively creating a comprehensive tapestry of NPO insights. This harmonious amalgamation formed the bedrock for subsequent analytical undertakings, imbuing the project with depth, context, and the potential for multifaceted exploration.

Now, the endeavor of wrangling scanned documents necessitated a markedly different journey, one that navigated through multiple layers of processing. The transformation of digitized documents, encapsulated within the format of PDF files, transformed these artifacts into Joint Photographic Experts Group (JPEG) images. This pivotal transformation was executed through the utilization of the *convert_from_path* module, harnessed from the *pdf2image* library. This conversion served as the cornerstone, laying the foundation for the subsequent infusion of OCR and NLP techniques.

3.1.2 Challenges

The images freshly obtained from the Candid repository, initially embedded within PDF files and subsequently transformed into the ubiquitous JPEG format, found themselves riddled by a series of challenges.

Foremost among these challenges was the imposition of noise, a disruptive element that cast an adverse influence upon the pristine quality of the newly converted images. This intrusion of noise stuffed in the images a certain degree of ambiguity, rendering them less clear and pristine than desired. A corollary challenge intertwined with this phenomenon manifested in the form of non-supportive contrast, an attribute that further worsened the impediments in visual clarity and interpretability.

An additional layer of complexity manifested in the realm of alignment. The process of manual scanning, undertaken by distinct individuals, engendered an element of non-uniformity in the alignment of the documents. This deviation from a standardized alignment precipitated potential challenges in character recognition. The non-conformity in alignment induced a distortion in the spatial integrity of the document, warranting a meticulous approach to ensure accurate character recognition.

The diversity of human involvement in the scanning process introduced another hurdle. The scans, due to the handling by distinct individuals. characterized by varied sizes and resolutions, thereby engendering an additional dimension of complexity. This variability in image dimensions and resolutions challenged the standardization efforts, necessitating a nuanced approach to harmonize and homogenize these images into a coherent corpus amenable to subsequent processing.

Moreover, since the nature of these forms was hand-written, the variations in strokes and styles of handwriting emerged as an inherent challenge. This divergence in penmanship introduced a layer of variability that demanded tailored interventions to ensure that the process of character recognition could effectively traverse these diverse expressions of handwriting.

In response to this intricate constellation of challenges, a suite of pre-processing techniques emerged formidable. These techniques, wielded strategically, aimed to enhance the precision of character recognition during the process of conversion. Among the preprocessing interventions, a critical approach was delineating the proximate vicinity of handwritten text that demanded conversion. This strategic demarcation sought to isolate the essential textual content from the ambient clutter, conferring a more focused and precise field for the conversion algorithms to operate within.

3.1.3 Setup

The environment used for this project was PyCharm. PyCharm offered an easy-to-use interface that made coding easier. Also debugging using this Integrated Development Environment (IDE) was much strategic given the availability of its robust debugging tools. The installation of this tool did not require much effort given its straight-forward and default settings. It also had a default environment path variable inclusion that reduces the effort of setting up in to half. Another major attribute of this IDE was that it offered a dark mode that was soothing to the eyes thereby allowing the prolonged exploitation of it.

3.1.4 Process

3.1.4.1 Data Preprocessing

The process of image enhancement and preparation ventured into an array of intricate methodologies, each strategically tailored to address specific challenges and enhance the overall quality of the images.

Binarization, realized through the meticulous application of Thresholding, emerged as a pivotal initial step. This process, rooted in the science of image processing, sought to transform the images into a binary form, where the foreground (textual content) and background could be unambiguously delineated. This operation laid the foundation for subsequent manipulations, setting the stage for heightened precision in character recognition and analysis.

Analogously, the employment of the *Gaussian Blur filter* assumed prominence as a strategic action to counteract the effects of noise, particularly the perturbing influence of salt-and-pepper noise and other minute artifacts. By applying this filter, the images underwent a process of smoothing, effectively diminishing the visual impact of undesirable noise and rendering the images more conducive to accurate character recognition. This intervention showcased the role of sophisticated image manipulation techniques in preparing the images for subsequent stages of processing.

In parallel, the application of *Histogram Equalization* emerged as a vital strategy to address the challenge of different contrast levels within the images. This technique, grounded in the principles of statistical analysis, facilitated the redistribution of intensity levels across the histogram of the image. The result was a more balanced distribution of intensity values, effectively enhancing the contrast and overall visual clarity. This endeavor encapsulated the fusion of mathematical concepts with image processing techniques to bolster the interpretability of the images.

In the context of text region identification and segmentation, the principle of *Connected Component Analysis* played a pivotal role. Through this method, the images were dissected into contiguous regions, effectively isolating distinct text areas and laying the groundwork for precise localization. The segmentation process, guided by connected components, facilitated a granular understanding of the textual content distribution within the images, which proved indispensable for subsequent analysis. This granularity was further harnessed through the mechanism of *Bounding Box Detection*. Employing this technique, bounding boxes were meticulously drawn around text regions, demarcating the boundaries between foreground text and the background. This step was instrumental in distinguishing the textual content from the surrounding elements, contributing to a more focused and accurate dataset that was conducive to subsequent analyses.

The process of resizing images carried with it the potential for distortion. To mitigate this risk, the application of a uniform aspect ratio was judiciously undertaken. By maintaining a consistent aspect ratio during the resizing process, distortions were effectively curtailed, ensuring that the intrinsic structure of the textual content remained faithful to its original form.

For images containing handwritten text, the *Stroke Width Transform* (SWT) emerged as a potent tool. This technique, rooted in computer vision principles, enabled the separation of text from the background by identifying the strokes within the handwritten content. Through the SWT, the strokes were isolated, effectively highlighting the textual content.

3.1.4.2 Data Extraction and Manipulation

The intricacies underlying the transformation of XML data into the Excel format within the Python programming paradigm necessitated a meticulously orchestrated association of key modules and libraries. This intricate process relied upon the interplay of the *xml.etree.ElementTree* module, intricately woven together with the versatile capabilities bestowed by the Workbook module sourced from the *openpyxl* library.

Central to this endeavor, the *xml.etree.ElementTree* module emerged as vital entity offering a profound mechanism for extracting crucial data from the XML source. With finesse, this module enabled the parsing of XML content, allowing for the extraction of vital information essential for the subsequent transformations. In parallel, the *openpyxl* library assumed a role of strategic importance, elevating the process through its adeptness at not only creating but also manipulating Excel spreadsheets.

The translation was inaugurated by fetching the pointer to its root tag. Meanwhile, the *openpyxl* library stepped forward, facilitating the inception of an Excel workbook. Within this canvas, data integration was achieved as information was transposed from the root to the leaf nodes my means of a seamless looping construct. This weaving of data into designated cells and sheets mirrored the inherent structure of the original XML content, culminating in an Excel representation that was both organized and comprehensive.

The upcoming stages would deal with the processing of the scanned hand-written files. The convergence of OCR and NLP methodologies was orchestrated on the preprocessed images as discussed before, with each technique contributing uniquely to the transformation. In the domain of OCR, the converted images underwent meticulous scrutiny, transmuting into their textual counterparts. The orchestration of this task was expertly conducted through a combination of the *Pillow* library and the *pytesseract* library. The former, a repository of image manipulation capabilities, collaborated seamlessly with the latter, an interface to the *Tesseract* OCR engine fortified by Google. The successful integration of the *Tesseract* OCR engine, ensconced within the local environment, empowers the utilization of its innate capabilities. This was made possible through the *image_to_string* function nested within the *Pillow* library; a tool used to successfully manipulate images.

The transition to the textual realm retained the spatial integrity and structural essence of the original content. This textual metamorphosis became the bedrock upon which subsequent processing progressed. Given the prowess of the *nltk* library, the textual document underwent tokenization carefully, resulting in its segmentation into discrete linguistic units guided by the inbuilt linguistic rules and heuristics of the library.

Based on specific conditional parameters, the target tokens, which represent distinct fields of information, were precisely located and extracted. Upon identifying a relevant token (field), the adjacent token was delicately extracted as the counterpart data of the field in question. This extraction, in terms of field-data pair, presented the extracted information in a comprehensive format.

The outcome resulting from this synergistic union between OCR and NLP is fool-proof. This synchronization enables the process with the capacity to extract specific and crucial data from scanned documents, including key attributes like organizational names and geographic location. This union stands as a verification to the potency of this approach in converting scanned documents into a decipherable and an analyzable format due to the ability to put Python libraries and sophisticated techniques to use.

3.2 Violation Detection Project

3.2.1 Data Collection

This segment of the project encompassed the utilization of two distinct sources for procuring images, namely images sourced from Google and images acquired through Geographic Information Systems (GIS). The collective image pool, which comprised more than 1500 images, served as the foundational dataset for constructing the model. The efficacy of this model was subsequently evaluated using a test set comprising approximately 200 images.

The pool of Google-sourced images was pulled from various third-party platforms, including notable entities like *Getty Images, Shutterstock* and *123rf*, alongside an assortment of miscellaneous sites affiliated with real estate insurance companies. Although a substantial portion of these images bore watermarks, their resolution was notably high. This characteristic conferred a heightened clarity upon the visual content, consequently facilitating the detection of violations within these images.

The acquisition of GIS images involved the strategic employment of specialized tools such as QGIS and ArcGIS. These applications were harnessed to gather image data from the Erie County of Buffalo [3]. However, it is important to note that the GIS-acquired images exhibited relatively lower resolutions, a factor that posed a noteworthy challenge in the context of violation identification. The principal obstacle entailed the accurate detection and demarcation of violations within these images. This preprocessing phase, necessitating the accurate identification of violations and their subsequent delineation, was paramount before integrating the images into the model during the training phase.

Central to this endeavor were 15 predominant violations that were accorded significant consideration during the construction of the model. These violations encompassed a spectrum of issues, such as **Bad Driveway, Bad Fencing, Bad Foundation, Bad Gutters, Bad Roof, Bad Siding, Bad Soffit, Bad Walkway, Broken Doors, Broken Windows, Damaged Garage Doors, Garage Debris, Peeling Paint, Sagging Porch, and Vehicles on Grass.** The images harnessed in this context were meticulously aligned to represent these violations, thereby establishing a comprehensive dataset that exemplified the diversity of potential infringements.

3.2.2 Challenges

The data collection phase of the Violation Detection project embarked on a journey fraught with multifold challenges, each contributing to the intricate tapestry of complexities that enveloped the project. As the trajectory of model construction unfolded, it became abundantly clear that the nascent challenges woven into the aggregated image dataset would significantly impact the efficacy of subsequent stages.

A palpable disparity emerged as a recurring theme, starkly apparent in pivotal attributes that govern image interpretation, including color, contrast, lighting, and clarity. This mélange of disparities cast a shadow on the feasibility of achieving a coherent level of generalization across images. The pronounced diversity was further exacerbated by the omnipresent watermark enigma that graced a substantial subset of images. Recognizing the weighty implications of these watermarks on accurate classification, the imperative for their meticulous removal assumed paramount importance. Failure to exorcise these watermark imprints carried the potential to skew the accuracy of the model, leading to biased classifications dictated by these extraneous elements.

The challenge posed by noise emerged as a formidable hurdle, with non-violation areas within images acting as harbors of confusion for the model. The injection of noise engendered a convolution of information, enshrouding the genuine signal under layers of interference. As such, the need to dissect signal from noise became an imperative, laying the groundwork for noise reduction strategies that aimed to cleanse the dataset and empower the model with more focused categorization capabilities.

Adding to the many layers of intricacies was the divergence in structural attributes across images. A plethora of variables unfurled, encompassing diverse shapes, varying sizes, and an array of orientations. This inherent variability rendered a standardized approach to model training a distant aspiration, demanding an adaptability that the model would need to inherently possess to navigate these complexities effectively.

Moreover, delving into the realm of Geographic Information Systems (GIS) data unfurled an entirely new chapter of challenges. While acquiring top-view images, critical for understanding roof configurations and yard layouts, was relatively within reach, the task of securing comprehensive side-view imagery emerged as an altogether distinct ordeal. The complexities inherent in obtaining these diverse perspectives further compounded the overarching challenges intrinsic to the data acquisition phase.

In response to these multifaceted challenges, strategic mitigation measures were devised. A comprehensive preprocessing regimen was introduced as the preliminary phase, meticulously designed to impose a semblance of uniformity upon the disparate dataset. The aim was twofold: to standardize attributes to enhance comparability and to combat noise that could potentially confound the learning process. However, it is crucial to acknowledge that certain challenges transcended the capabilities of the scope of this project, necessitating more specialized interventions and resources to achieve resolution.

3.2.3 Setup

The strategic selection of Google Colab as the development environment was a logical consequence of the availability of Graphics Processing Unit (GPU). This choice was underpinned by the understanding that the GPU computational prowess could be harnessed optimally within this platform, effectively capitalizing on its remarkable capabilities in handling image-intensive operations. The role of the GPU was particularly salient in mitigating the computational load associated with processing images.

3.2.4 Process

3.2.4.1 Data Processing

The execution was initialized by the implementation of preprocessing. Owing to the problems mentioned above, a string of techniques was used serially to enhance the quality of images at each step. This process was integral to ensuring that the images, regardless of their source, could be effectively harmonized for subsequent model training.

The Google-sourced images necessitated specific attention due to the presence of watermarks, which could adversely impact computer vision model accuracy by introducing irregularities that are then misconstrued as inherent elements of the images. As a countermeasure, advanced image processing techniques, including Image Inpainting and Deconvolution, were judiciously applied. However, it is noteworthy that this process, while eradicating irregularities, resulted in a marginal diminution of image quality. This trade-off, though, served an advantageous purpose as the real focus lay on the GIS images, which inherently did not boast the same degree of clarity even post-processing.

The GIS images underwent a distinct set of preprocessing interventions tailored to enhance their visual clarity and overall quality. A multi-pronged strategy was employed, encompassing alterations in the Red-Blue-Green (RGB) ratio, the application of *Gaussian filters*, manipulation of the standard deviation of an image, and the strategic utilization of *Gaussian pyramids*. These concerted efforts cumulatively elevated the image contrast and sharpness, thereby augmenting the visibility of the constituent elements. This enhancement was calibrated to strike a balance between human perception and computational discernment, acknowledging that the images needed to be human-readable while maintaining a semblance of recognizability for machine learning models.

3.2.4.2 Model Building and Selection

The models were meticulously crafted through a dual-phase strategy, commencing with the deployment of AWS *Rekognition*, followed by an intensive phase of architecting novel models that aimed to transcend the limitations of their predecessors.

In the AWS *Rekognition* phase, the foundational bedrock was laid by introducing the models to a dataset in its unaltered state, devoid of preprocessing. This dataset, comprising a diverse array of images, became a canvas upon which the intricacies of violations were vividly highlighted. This approach, manifested through meticulous highlighting techniques, enabled the model to glean a nuanced understanding of the precise spatial distribution and contextual attributes of these violations. The rationale behind this approach was deeply rooted in the objective of furnishing the model with an intuitive grasp of the targeted issues, a crucial step toward engendering a refined recognition capacity. Subsequently, the images were subjected to a rigorous preprocessing pipeline, refining their quality and stripping away artifacts and irregularities, thus culminating in a more refined dataset. This processed dataset, representative of a higher fidelity version of the original images, was reintroduced to the model. This iterative strategy underscored the commitment to holistic data refinement, imparting a layer of sophistication that elevated the predictive potential of the model.

However, the outcomes stemming from the AWS *Rekognition* phase revealed a performance that fell short of anticipated standards. This precipitated the commencement of the second phase, which revolved around the engineering of novel model architectures to surpass the limitations inherent in the AWS *Rekognition* paradigm. The need for a more dynamic and adaptable architecture prompted a shift towards the renowned capabilities of *PyTorch*, a framework uniquely poised to accommodate the creation of custom models from scratch.

Incorporating *PyTorch* as the backbone of this phase unleashed a spectrum of possibilities, allowing for the meticulous integration of customized architectural elements. The process of refining the models encompassed intricate permutations, with the integration of CAM and the strategic incorporation of *Residual Blocks*, thus representing an augmented form of the initial model architectures. The introduction of CAM offered an innovative dimension, endowing the model with the ability to discern and highlight specific regions within images that were crucial for accurate classification. This functionality not only bolstered the precision but also provided an insightful visualization into the decision-making mechanisms underpinning the predictions of the model. Since the *VGG-16* architecture lacks the residual blocks present in *ResNet-50* and to ensure an equal comparison between the two architectures, residual blocks were added. These blocks ensure faster learning pace towards the desired outcome.

The evaluation phase witnessed a robust comparative analysis between the *ResNet-50* and *VGG-16* models. The initial assessment revealed the superior performance of the *ResNet50* model. However, the pursuit of optimization persisted, culminating in a transformative refinement process for the *VGG-16* model. The infusion of residual blocks propelled the *VGG-16* model capabilities to a new echelon, effectively bridging the performance gap to a significant extent. Despite this noteworthy augmentation, the *ResNet-50* model sustained its lead in terms of performance.

As a culmination, five distinct models emerged from this comprehensive endeavor: **ResNet-50, VGG-16, VGG-16 with Residual Blocks, ResNet-50 with CAM, and VGG-16 with Residual Blocks and CAM. These models underwent an exhaustive evaluation process, encompassing a series of major model comparisons: ResNet-50 vs VGG-16, ResNet-50 vs VGG-16 with Residual Blocks, and ResNet-50 with CAM vs VGG-16 with Residual Blocks and CAM.**

Integral to this iterative process was the potency of *PyTorch,* which empowered the genesis of novel models from scratch. This flexibility facilitated not only the emulation of established architectures but also the dynamic integration of novel layers and structural adaptations. In a tapestry of meticulous craftsmanship, the journey from AWS *Rekognition* to bespoke model creation underlined the indomitable spirit of the undertaking in the quest for superior image data analysis.

4. Results and Discussion

The performance metrics for the Peoria project unfortunately remain unavailable due to the abrupt transition into the second project. As a result, a comprehensive assessment of the Peoria project outcomes and accomplishments could not be ascertained within the scope of this report.

Turning the spotlight onto the Violation Detection project, a realm of intriguing discoveries and revelations unfolded.

The pivotal yardsticks deployed to gauge the efficacy encompassed the **Precision, Recall, and the F1 score**, a comprehensive composite metric derived from the former two, synergistically furnishing a holistic evaluation of the performance of the models across a spectrum of classification scenarios. This trifecta of metrics emerged as an intricate apparatus to meticulously assess the competence in classifying images with varying degrees of precision and recall.

The inception of this evaluation journey transpired through the AWS *Rekognition* model, initially entrusted with unprocessed images. The results unveiled a disheartening scenario, with a precision value of 0.27 and a recall of 0.23, culminating in an F1 score of a mere 0.25. This outcome underscored the glaring insufficiencies of the model in achieving accurate and comprehensive classification. Notably, the implementation of image preprocessing heralded a marked improvement, bestowing a notable 20% surge in accuracy alongside enhanced precision, recall, and the F1 score. For this reinvigorated simulation, the precision ascended to 0.55, recall elevated to 0.49, and the F1 score scaled to 0.52. Despite this augmentation, the model performance remained unsatisfactory.

The quest for enhanced performance catalyzed the creation of custom CNN models, which infused a dynamic layer of adaptability. *The ResNet-50* model, synergistically woven with *VGG16* architecture, elicited precision and recall values of 0.611 and 0.57, consequently yielding an F1 score of 0.539. This competence was further amplified through the strategic assimilation of CAM into the model, resulting in a precision escalation to 0.642, an overarching recall of 0.721, and an F1 score of 0.607. This augmented iteration marked a perceptible leap from the AWS *Rekognition* phase, yet, the pursuit of refined performance persisted.

Subsequent efforts delved into the realm of standalone CNN models. The *ResNet-50* model, standing on its own, exhibited a recall of 0.611 and 0.57, thereby yielding an F1 score of 0.539. Meanwhile, the *VGG16* model, equipped with residual blocks, achieved a precision and recall of 0.45 and 0.52, respectively, contributing to an F1 score of 0.48. The strategic infusion of CAM into the latter resulted in a precision surge to 0.642, an overarching recall of 0.721, and an F1 score of 0.607. Notably, the *ResNet-50* model yielded a performance surpassing its counterparts, yielding a precision of 0.753, a recall of 0.816, and an F1 score of 0.78.

5. Conclusion and Outlook

Among the gamut of models evaluated, the *ResNet-50* model fortified with the strategic integration of CAM emerged as the pinnacle of performance, underscored by its superior precision, recall, and F1 score. Notably, this formidable model possesses the potential to serve as a foundational platform for subsequent iterations, offering a solid groundwork upon which further advancements can be meticulously crafted.

However, it remains unequivocal that the performance of the legacy model possessed untapped potential, awaiting refinement through enhanced image preprocessing methodologies and the potential incorporation of more sophisticated architectures. This prospective trajectory of improvement holds the promise of rendering the model even more adept at discerning violations within images, thereby realizing an elevated echelon of performance. This present model, although commendable, stands as a testament to the evolving nature of image classification algorithms and the perpetual pursuit of excellence.

An overarching constraint that exerted its presence throughout the project lifecycle was the paucity of a higher-caliber GPU. The constricted computational power endowed each simulation with a temporal burden, extending the completion time to a span of 3.5 to 5 hours. This confluence of limitations exacerbated the complexities inherent in fine-tuning the models, rendering the process laborious and protracted. The scarcity of computational resources cast a shadow over the efficiency of the optimization endeavor, necessitating meticulous planning and restraint in effecting changes.

Contemplating a parallel trajectory laden with more extensive research efforts and an augmented GPU resource evokes an outlook imbued with possibilities. With additional time and the availability of robust computational power, the prospects of conceiving a superior model are tantalizing. This envisaged model, inherently endowed with greater computational muscle, bears the promise of not only surpassing the current benchmark but also accelerating the violation detection process.

Note: The word count for the report text excluding the cover page, table of contents and references is **5278**

6. References

[1] https://www.irs.gov/charities-non-profits/tax-exempt-organization-search-bulk-data-downloads/
[2] https://www.guidestar.org/
[3] https://www3.erie.gov/gis/internet-mapping/