Highlights

Enabling New Interactions with Library Digital Collections: Automatic Gender Recognition in Historical Postcards via Deep Learning

- Users lack experience to navigate the unique materials hosted in online special collections, such as historical postcards.
- Special collections are increasingly releasing uncatalogued materials online, which provides even less metadata for users.
- Machine learning has emerged as a potential solution to derive additional metadata and support more detailed user queries.
- Prior research examined the use of text mining (e.g., for historical novels), but there is little work using computer vision.
- We create computer vision models that accurately extract gender from 28,308 postcards in special collections in 4.28 hours.

Enabling New Interactions with Library Digital Collections: Automatic Gender Recognition in Historical Postcards via Deep Learning

Abstract

The Walter Havighurst Special Collections from University Archives & Preservation at Miami University's King Library has a growing collection of over 600,000 historical postcards, with approximately 30,000 digitized, primarily from the Midwest during 1890-1919. This collection supports various lines of inquiry from users, such as analyzing the evolution of gender portrayal in popular media in the United States. However, manually separating the collection into postcards of males and females would take thousands of hours, which prevents the library from supporting sociological analyses at scale. Using an open postcard dataset, we trained deep neural networks to automatically detect people and classify them as male or female. We showed that this approach can accurately detect and classify females and confidently detect and label males for the library's collection of historical postcards. By employing deep neural networks, the library can enhance its metadata within hours and support new research inquiries at scale.

Keywords: Applied computer vision, Gender Recognition, Historical postcards, Object detection, Transfer learning

Enabling New Interactions with Library Digital Collections: Automatic Gender Recognition in Historical Postcards via Deep Learning

Ryan Schuerkamp[a], Jared Barrett[a], Amber Bales[b], Alia Wegner[b], Philippe J. Giabbanelli[a,*]

[a] *Department of Computer Science & Software Engineering,*
[b] *Miami University Libraries*
Miami University, 205 Benton Hall, Oxford, 45056-8819, Ohio, USA

Abstract

The Walter Havighurst Special Collections from University Archives & Preservation at Miami University's King Library has a growing collection of over 600,000 historical postcards, with approximately 30,000 digitized, primarily from the Midwest during 1890-1919. This collection supports various lines of inquiry from users, such as analyzing the evolution of gender portrayal in popular media in the United States. However, manually separating the collection into postcards of males and females would take thousands of hours, which prevents the library from supporting sociological analyses at scale. Using an open postcard dataset, we trained deep neural networks to automatically detect people and classify them as male or female. We showed that this approach can accurately detect and classify females and confidently detect and label males for the library's collection of historical postcards. By employing deep neural networks, the library can enhance its metadata within hours and support new research inquiries at scale.

Keywords: Applied computer vision, Gender Recognition, Historical postcards, Object detection, Transfer learning

*Corresponding author

Email addresses: `schuerr2@miamioh.edu` (Ryan Schuerkamp), `barret16@miamioh.edu` (Jared Barrett), `balesal2@miamioh.edu` (Amber Bales), `wegnera3@miamioh.edu` (Alia Wegner), `giabbapj@miamioh.edu` (Philippe J. Giabbanelli)

Acknowledgements: The authors are thankful for Davyd Sadovskyy, Stephen Zigmond, and Hongxiang He for their help annotating and filtering training and testing data.

Enabling New Interactions with Library Digital Collections: Automatic Gender Recognition in Historical Postcards via Deep Learning

Abstract

The Walter Havighurst Special Collections from University Archives & Preservation at Miami University's King Library has a growing collection of over 600,000 historical postcards, with approximately 30,000 digitized, primarily from the Midwest during 1890-1919. This collection supports various lines of inquiry from users, such as analyzing the evolution of gender portrayal in popular media in the United States. However, manually separating the collection into postcards of males and females would take thousands of hours, which prevents the library from supporting sociological analyses at scale. Using an open postcard dataset, we trained deep neural networks to automatically detect people and classify them as male or female. We showed that this approach can accurately detect and classify females and confidently detect and label males for the library's collection of historical postcards. By employing deep neural networks, the library can enhance its metadata within hours and support new research inquiries at scale.

Keywords: Applied computer vision, Gender Recognition, Historical postcards, Object detection, Transfer learning

1. Introduction

Digital libraries, digital collections, or digital archives all refer to supporting users of a university's library by providing online access to various materials. In the case of special collections, one (or more) platforms are often developed in addition to a library's main website in order to support users in navigating digitized heritage materials for teaching or research (Harden et al., 2022). However, as summarized by Burns et al. (2019), "when digitized and placed online, these unique materials are likely to attract a new and wider audience who may have little or no experience navigating special collections". Enabling access to special collections in ways that match the current expectations of users to obtain actionable data is thus a key challenge (Zavalina and Burke, 2023; Mulrennan, 2020). The problem is not necessarily resolved solely by giving and promoting a 'front end' online access to specialized collections (Mansbach, 2022). There is evidence regarding the 'back end' difficulty of keeping up with the item-level tasks for special collections at both well-endowed libraries and operations with smaller budgets (Pritchard, 2009).

Although archival specialists can assist users in navigating digital collections for specific research questions, human resources have limitations (Tammaro, 2020). Most importantly in this article, the number and type of records are important mediating factors. For example, the collection of 528,000 birth and death records at the University of Cincinnati[1] can be conveniently navigated online, as each record consists of a few fields (e.g., name, occupation, address) and several fields and searchable. In contrast, the collection of almost 11,000 historical ('dime') novels at Northern Illinois University[2] may lead to more open-ended research questions such as 'how were women historically portrayed' or 'has the rise of the contemporary environmental movement been paralleled with a growth in environmental themes in fiction'. Since users cannot query specific fields to answer these questions, and much less read every novel, they used machine learning to extract patterns from the text and increase the metadata available for queries (Short, 2019). There is thus a documented potential to integrate machine learning with the *discovery tools* that form part of the access technologies to special collections (Heyliger et al., 2016). The ability of machine learning algorithms to extract information may also contribute to alleviate the growing backlog problem of completely uncatalogued materials, particularly as special collections are making uncatalogued materials available online (Tam, 2017). In this paper, we explore the untapped potential of machine learning algorithms to automatically extract metadata from special collections of historical postcards, thus addressing both the backlog problem and supporting new lines of inquiries by potential users.

[1] `https://drc.libraries.uc.edu/handle/2374.UC/2032`

[2] `https://dimenovels.lib.niu.edu/`

2. Problem Statement

The Walter Havighurst Special Collections, University Archives & Preservation at Miami University's King Library has a physical collection of over 600,000 postcards from every state and grows as new donations are accepted. The digitization has been underway since 2013, with over 30,000 items currently digitized and catalogued (front picture, back picture, metadata such as whether it was mailed with a date/place of origin). Based on the 7,976 digitized postmarked instances, most of the digitized collection is from 1890-1919 (Figure 1). Providing an extensive set of metadata and supporting various lines of research on-demand for such a vast, growing, historical collection is an intensive process that vastly exceeds library resources. To address the need for automatic metadata extraction to assist both research needs and cataloging, we propose to use deep neural networks given their demonstrated strengths in analyzing images (Section 3). Specifically, we develop a machine learning method to automatically detect people in historical postcards and classify their gender as male or female[3].

Figure 1: The number of digitized postcards per decade.

3. Literature Review

3.1. Computer Vision for Historical Postcards

The applications of computer vision to historical postcards have been limited. Previous work has analyzed a collection of 1,346 postcards from World War I to classify postcards as one of eight categories: cartoon, destruction, frontline, landscape, love & poem, patriotism, portrait, or weapons (Grzeszick and Fink, 2014). Moreover, they demonstrated the difficulty of face detection on historical postcards and created a face detection method for historical images with a high precision of 94.7% but a low estimated recall of approximately 31% (Grzeszick and Fink, 2014). The authors also explored automatically grouping postcards by spotting potentially identical

[3]Given the historical and cultural context of the digitized collection, we classify gender as male or female, where gender refers to socially constructed roles, behaviors, and identities.

addressees (Fink et al., 2014). Additional work investigated layout extraction for historical postcards (García et al., 2022) using 100 digitized postcards from 1900-1930 to assign each pixel to one of seven categories in the postcard image layout: background, postmark, stamp, handwritten text, printed text, auxiliary lines, and image or drawing.

3.2. Object Detection

Object detection is a subfield of computer vision that determines where and what objects occur in an image. Object detection algorithms predict *bounding boxes* enclosing desired objects in an image and classes for each bounding box. To train an object detection model, training images are manually annotated with bounding boxes enclosing items of interest with labels indicating the type of the object. For example, to detect people and their gender, people in training images would be enclosed by bounding boxes with their gender labeling the box (Figure 2). Object detection models use these bounding boxes for reference during training to learn what defines the labeled item within the bounding box.

Figure 2: Bounding boxes enclose people. A blue bounding box indicates a person is a male, and a red bounding box denotes a female.

Two popular approaches for object detection are Region-Based Convolutional Neural Networks (R-CNN) (Girshick et al., 2014) and You Only Look Once (YOLO) (Redmon et al., 2016). R-CNN is a two-step process: it first predicts bounding boxes, then classifies their content (Girshick et al., 2014). Numerous extensions have been proposed to improve accuracy and decrease inference time (Girshick, 2015; Ren et al., 2015), but its two-stage approach can still hinder its inference speed. In contrast, the core idea of YOLO is "you only look once": the algorithm draws bounding boxes and detects objects simultaneously, resulting in a faster inference speed (Redmon et al., 2016). Since the initial proposal, several modifications have also sought to increase accuracy (Jiang et al., 2022). Both approaches leverage convolutional neural networks (Gu et al., 2018), which are widely used in computer vision (Li et al., 2022).

3.3. Measuring Performance via Cross-validation

In machine learning and computer vision, a dataset is split into a training set used to build and select models and a testing set used to evaluate the performance of the final model. If this split occurred only once, then the model's performance may partially be attributable to luck, and parameters of the machine learning algorithm may not be optimized. To provide a rigorous evaluation and perform an optimization process, the training set is also split, thus resulting into training/validation/testing sets. Hyper-parameters of the machine learning process are optimized on these splits, then the model is trained with the best values using the entire training set and finally applied onto the testing set. This process is called *cross-validation* (Berrar, 2019) and each of the splits results in non-overlapping *folkds* with approximately the same number of instances. This is a widely used process used in machine learning (King et al., 2021) and computer vision (DeCost and Holm, 2015), including object detection (Gonzales-Martínez et al., 2021).

For a classification problem, it is critical to ensure a balanced number of classes in each fold; otherwise, the model may favor the majority class (Japkowicz and Stephen, 2002). For example, if there are 50 men and 50 women and 10 folds, it is ideal to have 10 folds each with five men and women. This balanced k-fold cross-validation is called *stratified k-fold cross-validation*. In some applications such as ours, each observation may contain instances of multiple classes (e.g., an image can have three women and one man), so stratification with multiple labels is necessary. If multiple labels are balanced with k-fold cross-validation, the process is called *multi-label stratified k-fold cross-validation* (Sechidis et al., 2011).

3.4. Data Augmentation

Generalizability, the ability to apply the model to never-before-seen data, is a challenge in machine learning and computer vision. Small sample sizes cannot represent many variations, and it can be difficult to get a large number of samples (Shorten and Khoshgoftaar, 2019), especially in certain application areas like medical imaging and when obtaining and annotating training data is expensive for a task such as object detection (Kaur et al., 2021). Without enough training data, overfitting and large class imbalances can occur, impairing the model's accuracy. Therefore, the quantity and diversity of training data are critical to the model's accuracy. Data augmentation is the process of using existing data to generate new data points and is a popular approach used to increase the quality and size of the training set. It is widely used in computer vision (Shorten and Khoshgoftaar, 2019; Kaur et al., 2021; Wang et al., 2020), including in sensitive domains like medical imaging (Chlap et al., 2021; Nalepa et al., 2019), to increase the size of the training data, correct class imbalances, and produce models against common alterations (e.g., slightly rotated images).

In computer vision, several data augmentation techniques can help make a model robust to common distortions and significantly increase its accuracy (Kaur et al., 2021). For example, randomly removing part of an image (Figure 3) can force the model to search for other relevant content while the most discriminative content is hidden, reducing the risk of over-fitting, making the model robust to occlusion, and boosting model performance on the given task (Zhong et al., 2020). Other common techniques include horizontally flipping the image and randomly cropping or rotating the image (Figure 3), which can make the model robust to positional bias. Additionally, Kernel filters, like gaussian blur, increase accuracy on blurred images (Shorten and Khoshgoftaar, 2019).

There are three main challenges with data augmentation for computer vision. First, there is no universal way to select suitable augmentations; it depends on the problem and dataset. For example, vertical flipping images in facial recognition may not increase performance because vertically flipped images are likely rare in the testing set and real-world data. Second, some data augmentations may not preserve the labels or items of interest in the image (Shorten and Khoshgoftaar, 2019). For instance, image translation may move bounding boxes out of the image, and rotation may cause a digit recognition model to mistake a 9 for a 6 (Shorten and Khoshgoftaar, 2019). Third, data augmentations can reduce model performance instead of improving it, as exemplified in studies on random cropping (Yang et al., 2021). As a result, the selection of data augmentation techniques is important and relies on both an understanding of the data and computer vision algorithms.

Figure 3: Five common data augmentations.

3.5. Transfer Learning

Transfer learning attempts to improve performance in one target domain by leveraging information from a related domain, similar to how a person with experience playing guitar can use that knowledge to learn to play the piano more effectively than someone without a musical background (Weiss et al., 2016). Transfer learning is frequently used in vision tasks such as object detection (Talukdar et al., 2018) because training data in the target domain is rare, expensive to collect and label, or inaccessible (Weiss et al., 2016). Transfer learning is achieved by freezing some layers of the neural network, which preserves knowledge from the previous domain (Jocher et al., 2022) and allows to adapt to a new domain with minimal costs for re-training. For example, freezing all but the output layer prevents any of the frozen layers' weights from changing but allows the output layer's weights to change in order to predict new classes (Jocher et al., 2022). Additionally, training may start from pre-trained weights, which were trained on a dataset in a different domain to leverage knowledge from that domain (Jocher et al., 2022).

4. Methodology

Figure 4: Our methodology.

We followed a six-step process to predict the number of males and females in historical postcards (Figure 4):

1. We *created* a training set from postcards available online, in order to create a reusable model that can be applied to various historical collections. The postcards obtained from our library were held separately as a testing set.
2. We *balanced* the number of males and females using multi-label stratified 10-fold cross-validation (Section 3.3), thus avoiding the creation of a model that is biased in performing well on one type at the expense of the other.
3. We *augmented* the data by applying seven methods (Section 3.4) independently to each image in the training set. This increase in size and variety of the training set ultimately contributes to the robustness of the model.
4. We *trained* YOLOv5x object detection models (Jocher et al., 2022) through transfer learning (Section 3.5) starting from the pre-trained weights, which were trained on the COCO dataset (Lin et al., 2014).
5. We *evaluated* the models with the best hyper-parameter combinations on the validation set, originating from online postcards. This provides insight into the ability of the models to accurately detect and classify males and females for postcards that are similar to the training data.

6. We *tested* the best-trained model on postcards from the library's digitized collection, thus evaluating the ability of our model to generalize beyond their training set and be readily used by a library.

All programming used Python 3.7.1 with several packages, listed in Table 1. Our code is publicly and permanently available on an anonymized repository at `doi.org/10.5281/zenodo.7689513`. Each of the six consecutive steps is detailed in a subsection below.

4.1. Build Training and Testing Set

The library provided 200 images from their historical postcard collection for testing. A total of 2,028 additional images were gathered from various repositories to create a training set. Once the data was collected, we dispatched each image to an annotator who used CVAT (Sekachev et al., 2020) to enclose each person by a bounding box and choose a male or female label. We employed five annotators and ensured consistency in bounding boxes and labels as the work of each annotator was reviewed by another. Images were eliminated for any one of four reasons (Figure 5). After elimination, images in the final testing set were assigned a time period using the date postmarked or the input of two subject matter experts.

Figure 5: The four reasons postcards were removed from the testing set.

4.2. Multi-label Stratified 10-Fold Cross-validation

We performed multi-label stratified 10-fold cross-validation (Sechidis et al., 2011) to create 10 training and validation sets to obtain the best hyperparameter combination for each model; the training and type of models are described in section 4.4. We leveraged stratification to approximately balance the number of males and females in each training and validation set and avoid skewing the predictions towards the majority class (Japkowicz and Stephen, 2002).

4.3. Data Augmentation

Each image in the training set went through seven augmentation techniques (Figure 6) to increase the size of the training dataset, avoid overfitting (Shorten and Khoshgoftaar, 2019), and make the model robust to positional bias (e.g., a person appearing in a different position in a postcard) (Yang et al., 2021) and other common variations to increase model accuracy (Shorten and Khoshgoftaar, 2019). We chose to shear and equalize each postcard because shearing and equalizing are frequently used in optimal augmentation policies (i.e., sequences of augmentations) for object detection (Zoph et al., 2020). Shearing distorts the position and size of the bounding box, whereas equalizing improves the contrast and does not modify the bounding boxes (Zoph et al., 2020). We selected random cropping and random horizontal flipping (i.e., flipping the image with a certain probability) because they are the standard configuration for object detection tasks with limited training data (Yang et al., 2021), and random cropping may help increase the spatial robustness of the model since the object of interest may appear in different locations (Yang et al., 2021). Additionally, horizontally flipping the image has increased performance on popular computer vision datasets such as ImageNet (Shorten and Khoshgoftaar, 2019). Note, we did not select two popular augmentations (rotation and vertical translation) because they resulted in bounding boxes extending out of the image (i.e., people would be moved off screen).

To further increase the quantity and diversity of training data, we also created three additional augmentation strategies by applying combinations of some of the four augmentations discussed above. A combination consists of applying several augmentations, one after the other. This is known as a 'sequential augmentation' and it is a common policy in computer vision to improve performances on object detection tasks (Zoph et al., 2020). The first policy horizontally flips and equalizes the image. The second policy starts by shearing the image, and the third starts by randomly cropping the image. Then, the second and third policies horizontally flip the image half the time and equalize it.

Figure 6: A postcard and the seven additional training images produced by each augmentation.

4.4. Training

We trained six versions of the YOLOv5x model (Jocher et al., 2022) by creating variants where none of the layers was frozen (Unfrozen), the backbone was frozen (Backbone), and all but the last output layer frozen (Last Layer). Models were trained on the augmented and unaugmented version of each of the 10 folds starting from the pre-trained YOLOv5x weights trained on the COCO dataset (Lin et al., 2014) to detect and classify people as male or female. As a result, 60 models were trained in total. Starting from the pre-trained weights usually reduces training time (Jocher et al., 2022), and the COCO dataset contains people, so the model can leverage the domain knowledge of detecting people and may only need minor modifications to predict gender. We selected a YOLO-based model because of YOLO models' strong performance on object detection benchmarks such as COCO (Lin et al., 2014) and PASCAL VOC (Everingham et al., 2012) and their ability to use additional features instead of just facial features to detect people, whereas many popular approaches for gender detection rely solely on facial features (Azzopardi et al., 2018; Antipov et al., 2017; Dhomne et al., 2018) and only perform well on high-resolution images free of corruptions (Greco et al., 2021), degrading their performance on historical postcards, especially when people's faces are not fully visible. The models were trained with a Tesla V100-PCIE-16GB GPU for 300 epochs with a batch size of 8, an image size of 640, and 1 worker.

4.5. Validation and Testing

We performed validation and testing in a six-step process. First, we applied each model to the validation set for its fold after training. The trained models may predict someone as both male and female, so there may be two bounding boxes around a person, each with their own confidence (Figure 7). Our second step eliminated these overlapping bounding boxes. We used the *intersection over the union* (IoU) (Figure 8) to determine how much two bounding boxes overlap, and we introduced a parameter to detect whether two bounding boxes (one predicting male and the other female) enclose the same person. We varied the parameter in the interval [0.01, 1] in increments of 1, where 1 means the bounding boxes must perfectly intersect to be considered the same, and 0.01 considers two bounding boxes as the same if they intersect by at least 1% (Figure 9). If the IoU of two bounding boxes meets or exceeds the threshold, we select the bounding box and prediction with the higher confidence.

Figure 7: Overlapping male and female predictions.

Third, we assessed the accuracy of the predictions with the labeled ground-truth validation set for each of the 10 folds. While two bounding boxes could

Figure 8: The intersection over the union (IoU) is calculated by dividing the area of intersection by the area of the union, producing a value between 0 (no overlap) and 1 (complete overlap).

Figure 9: Example intersection over the union (IoU) values. The higher the value, the more the bounding boxes overlap.

occasionally be produced in step two, it becomes much more frequent in step three since we compare predicted bounding boxes with ground-truth bounding boxes. We thus use the same strategy of measuring the IoU and applying a parameter to determine whether two boxes are the same. We varied the parameter in the interval [0.5, 1] in increments of 0.01, where a value of 0.5 means the bounding boxes must intersect by at least 50% to consider them a match. A minimum value of 0.5 is in-line with widely-used object detection datasets such as COCO (Lin et al., 2014) and PASCAL VOC (Everingham et al., 2012). For each predicted bounding box, we checked if it matches a ground-truth bounding box. If so, we then check if the classes (male and female) also match. For completeness, we also check if bounding boxes from the ground-truth dataset were missed by the model's predictions.

Fourth, we optimized the two IoU parameters (to determine overlapping predictions or a match with the ground-truth) for accuracy and F1. This is part of finding the best combinations of hyper-parameter values to detect and classify males and females. Fifth, as expected in a nested cross-fold validation, we applied the hyper-parameter combinations with the highest accuracy and F1 found in the validation set onto the testing set for each of the 60 trained models. Finally, we calculated the accuracy, precision, recall, and F1 for each model's best hyper-parameter combinations for accuracy and F1 on the testing set. We calculated these metrics for the overall testing set and also for the subset of postcards from 1890-1919, since most of library's collection is from this time period.

4.6. Inference on Collection

The trained model with the highest performance on the testing set was applied to the front image of 28,308 digitized postcards, predicting the number of males and females in each postcard. Inference was performed with a Tesla V100-PCIE-16GB GPU.

5. Findings

5.1. Data Pre-Processing: Balancing and Augmentation

After annotating and eliminating postcards, we created a balanced *training set* of 987 images with 1238 males and 1238 females. The testing set consisting of 123 images with 150 and 160 males (Figure 10), including 72 images with 96 females and 115 males from 1890-1919. Performing multi-label stratified 10-fold cross-validation produced 10 training and validation sets with approximately the same number of males, females, and images (Table 2). We augmented each fold's training set (Table 3), generating 5677 augmented images and bringing the size of the training set to 6664 images.

Figure 10: Breakdown of the 77 postcards removed from the testing set.

5.2. Validation and Testing

The best of the 60 trained models have accuracy, precision, recall, and F1 over 0.9 for both males and females on their validation sets after optimizing the IoU duplicate threshold and IoU ground-truth threshold for accuracy (Tables 4 and 5) and F1 (Tables 6 and 7). Note, different values of the IoU duplicate threshold and IoU ground-truth threshold hyper-parameters vastly affect the model's F1 and accuracy (Figure 11). Augmented Backbone Fold 4, the model trained with its backbone frozen on the augmented version of fold 4, has the best performance according to accuracy and F1 for females overall and from 1890-1919 on the testing set. From 1890-1919, it has an accuracy of 0.69182 (Table 8), and it has an overall accuracy of 0.60265 and precision of 0.69841 (Table 9). Unaugmented Backbone Fold 3, the model trained with its backbone frozen on the unaugmented version of fold 3, has the best performance according to accuracy and F1 for males overall and from 1890-1919 on the testing set. From 1890-1919, it has an accuracy of 0.58824 and a precision of 0.94872 (Table 10), meaning when the model predicts a male, it is a male 94.872% of the time. Overall, it has an accuracy of 0.63844 and a precision of 0.92727 (Table 11). For the sake of brevity, we omit the tables of the best models according to F1 because the best performing models are the same as the ones for accuracy. Our full results for all 60 models on the validation and testing sets are available in a permanent and anonymous repository at `doi.org/10.5281/zenodo.7689882`. Unaugmented Backbone Fold 3 completed inference on 28,308 postcards in 4.28 hours, finding 11,313 females and 12,164 males.

Figure 11: The effects of the IoU duplicate threshold and IoU ground-truth threshold on model F1 and accuracy for a model trained with its none of its layers frozen on the augmented version of fold 2 (i.e., Augmented Unfrozen Fold 2). Note, the F1 is undefined when the IoU ground-truth threshold is 1, producing the blank values in that column.

6. Discussion

Given the difficulty at navigating unstructured data in a library's special collections, and the growing trend of making uncatalogued materials available online, there is a documented need to efficiently extract metadata from such collections at scale. Metadata can contribute to cataloging efforts and provides new data fields that support user queries. In this paper, we focused on a historical postcard collection which faces both of the challenges evoked: data is not fully catalogued (10 years have allowed to cover 5% of the data) and would not support user queries related to gender. Studies of gender portrayal in historical collections are popular library projects for students (Pankuch and Wilson, 2019; Teaching et al., 2022) and an active subject of inquiry at the crossroad of history and identity (Ringer, 2007; Anagnostopoulous, 2011; Chakraborty, 2018). For example, the subtle gendered shift in occupations and traditions shown in postcards can contribute to understand broader trends in national identities and social transformations (Pite, 2021). By using machine learning techniques from computer vision, our study shows that gender metadata can accurately be added to historical postcard collections, thus supporting scholarship and the missions of libraries' special collections.

The trained models successfully achieved our goal of detecting gender in historical postcards, especially in the Midwest between 1890-1919. Our best models for this time period had an accuracy of 0.69182 on detecting and classifying females and precision of 0.94872 on detecting and classifying males. While human accuracy would be higher, our model was able to process 28,308 historical postcards within 4.28 hours, while no human may even want to process such a volume of records. Together, these results enable researchers to use new filters and quickly search a special collection for images with a target gender.

Our precision of 94.872% and recall of 33.036% from 1890-1919 on male gender detection exceed the precision of 94.7% and the recall of approximately 31% for face recognition on World War I postcards in past studies (Grzeszick and Fink, 2014). This improvement in results is partly attributable to accounting for a person's entire body, as our bounding box is not limited to faces. Our results also demonstrate the viability of a novel application of computer vision to library sciences and historical postcards on a collection much larger than the 1,346 (Grzeszick and Fink, 2014) and 100 (García et al., 2022) postcards analyzed by previous methods. Furthermore, the performances of our 'augmented backbone' model demonstrate a successful application of data augmentation to historical postcards, which was not leveraged in previous applications of computer vision to historical postcards.

The collection primarily has postcards from the Midwest from 1890-1919, and our analysis and model building focused on building the best possible model for the collection. As a result, our models may not perform as well on images from before 1890, after 1919, outside the Midwest or United States, or that are not postcards. For example, our models' performance may suffer when applied in different cultural contexts where men may also wear longer clothing (Figure ??), since such garments were primarily worn by women within the context of the Midwest from 1890-1919. As a result, it would be of particular interest to create historically and culturally aware models capable of accurately detecting people and classifying their gender in different environments (Preston, 2009). Broadening the capacity of our models would support studies in visual anthropology, particularly for marginalized individuals (Cheung, 2000).

7. Conclusion

We aimed to develop an automatic method to detect people and classify their gender in historical postcards to enable a more complex analysis of a growing collection of over 600,000 (30,000 digitized) historical postcards and the evolution of gender portrayal in popular media in the United States. We successfully trained models capable of accurately identifying females and confidently labeling males in historical postcards, especially in the Midwest from 1890-1919, while demonstrating the immense time savings of applying our models to the collection over manual analysis. Our work exhibits a novel application of computer vision for historical postcards and the power of machine learning to quickly and efficiently analyze vast amounts of data, supporting researchers in further sophisticated analysis.

References

Anagnostopoulous, A. (2011). Picturing public space: Ethnicity and gender in picture postcards of iraklio, crete. *Audiovisual Media and Identity Issues in Southeastern Europe*, page 171.

Antipov, G., Baccouche, M., Berrani, S.-A., and Dugelay, J.-L. (2017). Effective training of convolutional neural networks for face-based gender and age prediction. *Pattern Recognition*, 72:15–26.

Azzopardi, G., Greco, A., Saggese, A., and Vento, M. (2018). Fusion of domain-specific and trainable features for gender recognition from face images. *IEEE Access*, 6:24171–24183.

Berrar, D. (2019). Cross-validation.

Burns, D., Crane, B., Pumphrey, D., and Thoms, B. (2019). What we talk about when we talk about digital libraries: Ux approaches to labeling online special collections. *Weave: Journal of Library User Experience*, 2(1).

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Albumentations: fast and flexible image augmentations. *Information*, 11(2):125.

Chakraborty, S. (2018). Mammies, ayahs, baboes: Postcards of racialized nursemaids from the early twentieth century. *Visual Culture & Gender*, 13:17–31.

Cheung, S. C. (2000). Men, women and "japanese" as outsiders: A case study of postcards with ainu images. *Visual Anthropology*, 13(3):227–255.

Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., and Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563.

Clark, A. (2015). Pillow (pil fork) documentation.

DeCost, B. L. and Holm, E. A. (2015). A computer vision approach for automated analysis and classification of microstructural image data. *Computational Materials Science*, 110:126–133.

Dhomne, A., Kumar, R., and Bhan, V. (2018). Gender recognition through face using deep learning. *Procedia Computer Science*, 132:2–10. International Conference on Computational Intelligence and Data Science.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2012). The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

Fink, G. A., Rothacker, L., and Grzeszick, R. (2014). Grouping historical postcards using query-by-example word spotting. In *2014 14th International Conference on Frontiers in Handwriting Recognition*, pages 470–475.

García, B., Moreno, B., Vélez, J. F., and Sánchez, Á. (2022). Deep layout extraction applied to historical postcards. In Ferrández Vicente, J. M., Álvarez-Sánchez, J. R., de la Paz López, F., and Adeli, H., editors, *Bio-inspired Systems and Applications: from Robotics to Ambient Intelligence*, pages 346–355, Cham. Springer International Publishing.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Gonzales-Martínez, R., Machacuay, J., Rotta, P., and Chinguel, C. (2021). Faster r-cnn with a cross-validation approach to object detection in radar images. In *2021 IEEE International Conference on Aerospace and Signal Processing (INCAS)*, pages 01–04.

Greco, A., Saggese, A., Vento, M., and Vigilante, V. (2021). Gender recognition in the wild: a robustness evaluation over corrupted images. *Journal of Ambient Intelligence and Humanized Computing*, 12:10461–10472.

Grzeszick, R. and Fink, G. A. (2014). Recognizing scene categories of historical postcards. In Jiang, X., Hornegger, J., and Koch, R., editors, *Pattern Recognition*, pages 604–615, Cham. Springer International Publishing.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., and Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377.

Harden, M., Martinez-Montavon, A. M., and Narlock, M. (2022). A mismatched group of items that i would not find particularly interesting: Challenges and opportunities with digital exhibits and collections labels. *Evidence Based Library and Information Practice*, 17(4):71–121.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Heyliger, S., McLoone, J., and Thomas, N. L. (2016). Moving toward "megachoice": The evolution of access technologies in special collections. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage*, 17(1):17–31.

Japkowicz, N. and Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449.

Jiang, P., Ergu, D., Liu, F., Cai, Y., and Ma, B. (2022). A review of yolo algorithm developments. *Procedia Computer Science*, 199:1066–1073. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 & 2021): Developing Global Digital Economy after COVID-19.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., TaoXie, Michael, K., Fang, J., imyhxy, Lorna, Wong, C., Yifu, Z., V, A., Montes, D., Wang, Z., Fati, C., Nadar, J., Laughing, UnglvKitDe, tkianai, yxNONG, Skalski, P., Hogan, A., Strobel, M., Jain, M., Mammana, L., and xylieong (2022). ultralytics/yolov5: v6.2 - YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai integrations.

Kaur, P., Khehra, B. S., and Mavi, E. B. S. (2021). Data augmentation for object detection: A review. In *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 537–543.

King, R. D., Orhobor, O. I., and Taylor, C. C. (2021). Cross-validation is safe to use. *Nature Machine Intelligence*, 3(4):276–276.

Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2022). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer.

Mansbach, S. (2022). Historical collections for the future: A note on modernist books at columbia and cornell. *Slavic & East European Information Resources*, 23(1-2):240–250.

Mulrennan, K. (2020). Lessons in making the unique ubiquitous: diversifying the role of the special collections and archives department to enhance teaching and learning at the university of limerick. *Archives and Records*, 41(2):126–147.

Nalepa, J., Marcinkiewicz, M., and Kawulok, M. (2019). Data augmentation for brain-tumor segmentation: a review. *Frontiers in computational neuroscience*, 13:83.

Pankuch, A. and Wilson, J. (2019). "feminists leap year vol. 2:" the portrayals of gender in early 20th century postcards. *Student Projects from the Archives*, 1(1):2.

Pite, R. E. (2021). The rural woman enters the frame: A visual history of gender, nation, and the goodbye mate in the postcolonial río de la plata. *Journal of Social History*, 54(4):1120–1159.

Preston, R. (2009). 'hope you will be able to recognise us': the representation of women and gardens in early twentieth-century british domestic 'real photo'postcards. *Women's History Review*, 18(5):781–800.

Pritchard, S. M. (2009). Special collections surge to the fore. *portal: Libraries and the Academy*, 9(2):177–180.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Ringer, G. (2007). Gender posed: the people behind the postcards. In *Tourism and gender: Embodiment, sensuality and experience*, pages 219–234. CABI Wallingford UK.

Sechidis, K., Tsoumakas, G., and Vlahavas, I. (2011). On the stratification of multi-label data. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III 22*, pages 145–158. Springer.

Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., TOsmanov, Kruchinin, D., Zankevich, A., DmitriySidnev, Markelov, M., Johannes222, Chenuet, M., a andre, telenachos, Melnikov, A., Kim, J., Ilouz, L., Glazov, N., Priya4607, Tehrani, R., Jeong, S., Skubriev, V., Yonekura, S., vugia truong, zliang7, lizhming, and Truong, T. (2020). opencv/cvat: v1.1.0.

Short, M. (2019). Text mining and subject analysis for fiction; or, using machine learning and information extraction to assign subject headings to dime novels. *Cataloging & Classification Quarterly*, 57(5):315–336.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48.

Talukdar, J., Gupta, S., Rajpura, P. S., and Hegde, R. S. (2018). Transfer learning for object detection using state-of-the-art deep neural networks. In *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 78–83.

Tam, M. (2017). Improving access and “unhiding” the special collections. *The Serials Librarian*, 73(2):179–185.

Tammaro, A. M. (2020). Covid 19 and libraries in italy. *International Information & Library Review*, 52(3):216–220.

Teaching, I. H. E., Association, L., and Wissner, R. A. (2022). Active images: Teaching students critical thinking and analysis skills with picture postcards.

Wang, X., Wang, K., and Lian, S. (2020). A survey on face data augmentation for the training of deep neural networks. *Neural Computing and Applications*, 32(19):15503–15531.

Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, 3(1):9.

Wes McKinney (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.

Yang, R., Wang, R., Deng, Y., Jia, X., and Zhang, H. (2021). Rethinking the random cropping data augmentation method used in the training of cnn-based sar image ship detector. *Remote Sensing*, 13(1).

Zavalina, O. L. and Burke, M. (2023). User searching in digital repositories. In *Discoverability in Digital Repositories*, pages 88–114. Routledge.

Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y. (2020). Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13001–13008.

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T.-Y., Shlens, J., and Le, Q. V. (2020). Learning data augmentation strategies for object detection. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020*, pages 566–583, Cham. Springer International Publishing.

Figure 1: The number of digitized postcards per decade.

Click here to access/download;Figure;Figure 1.png

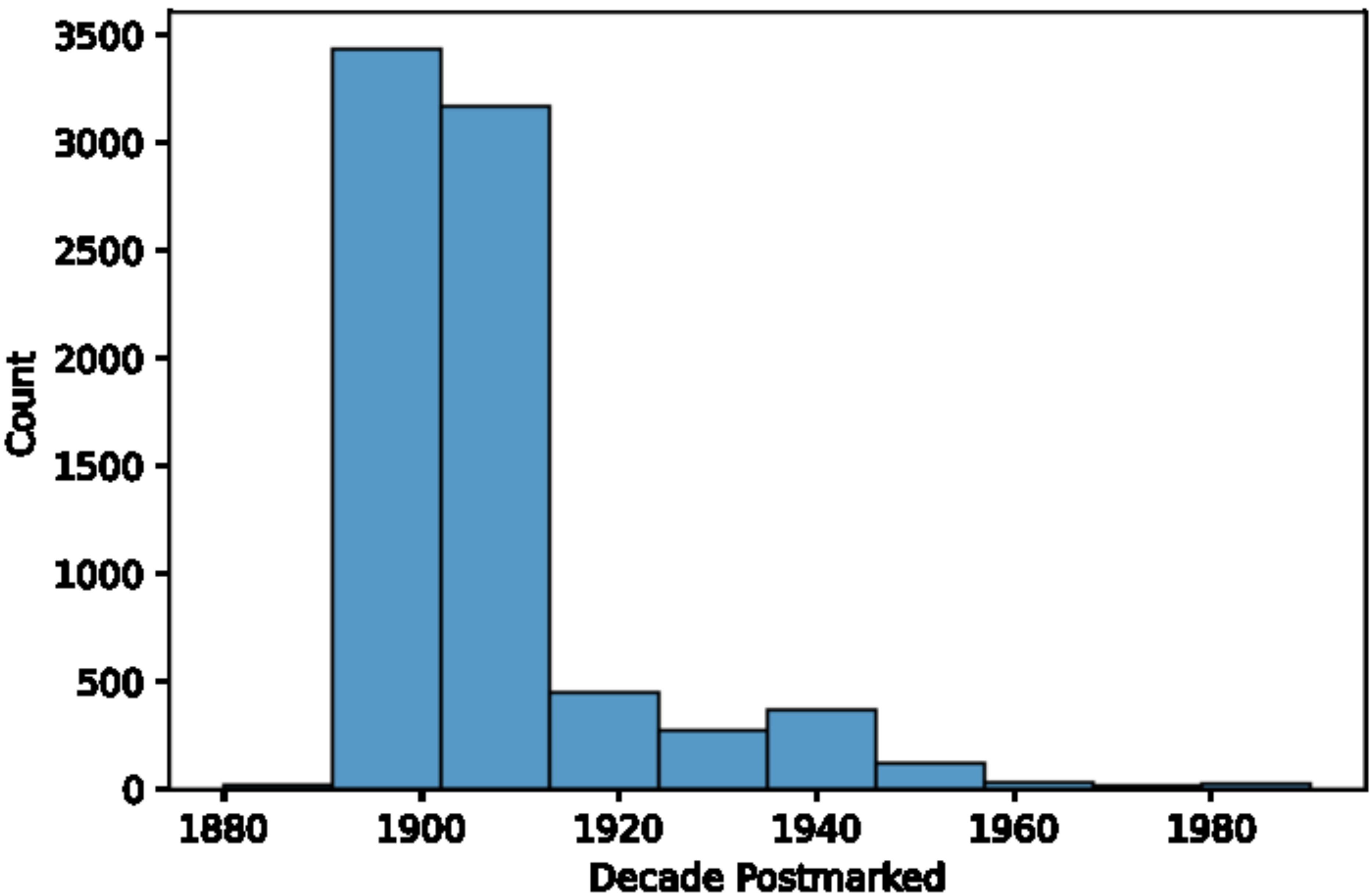


Figure 2: Bounding boxes enclose people. A blue bounding box indicates a person is a male, and a red bounding box denotes a female

Figure 3: Five common data augmentations.

Click here to access/download;Figure;Figure 3.png

Figure 4: Our methodology

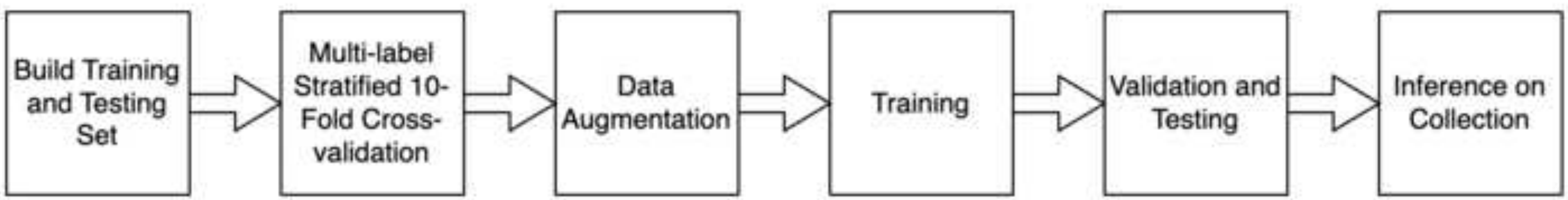


Figure 5(a) In some postcards, humans cannot tell how many people there are.

Click here to access/download;Figure;Figure 5a.png

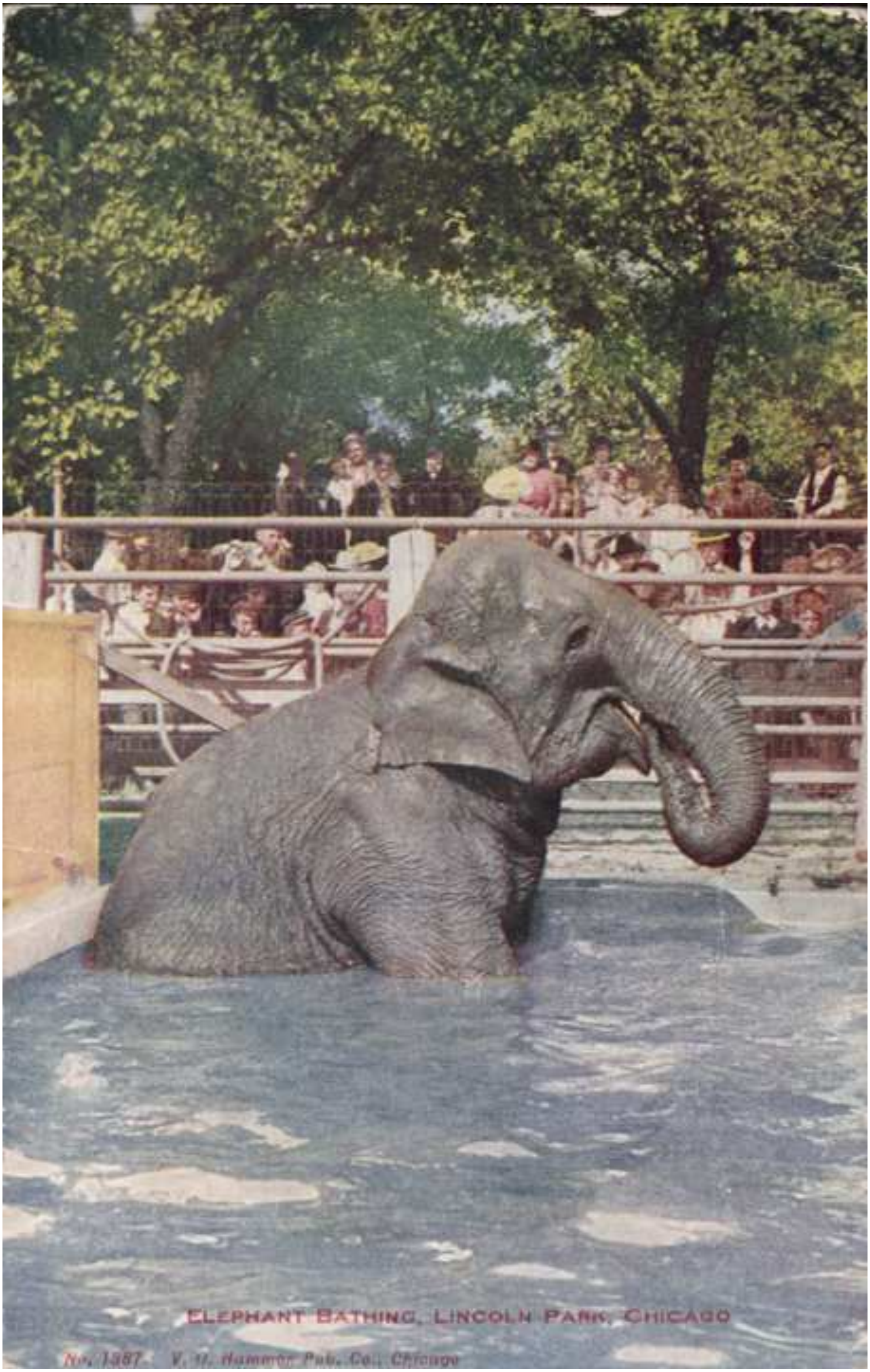


Figure 5(b) Young boys and girls would wear dresses in the early 20th century, making it difficult to determine gender.

Click here to access/download;Figure;Figure 5b.png

Figure 5(c) Some postcards were paintings or pictures of sculptures instead of pictures of people.

Click here to access/download;Figure;Figure 5c.png

Figure 5(d) The vast majority of the collection is from the Midwest, so postcards from the middle east are not relevant for our model.

Click here to access/download;Figure;Figure 5d.png

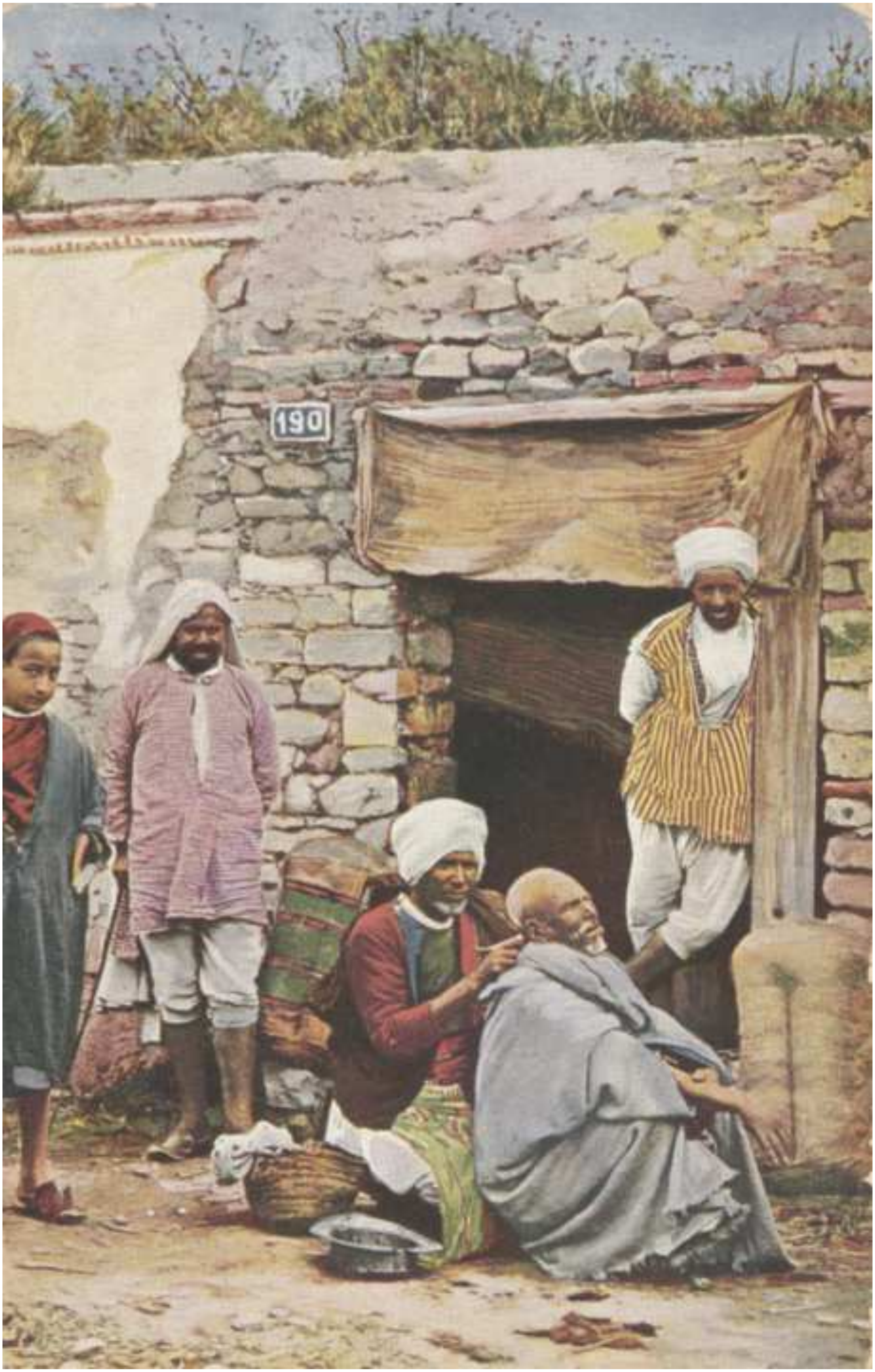


Figure 6: A postcard and the seven additional training images produced by each augmentation.

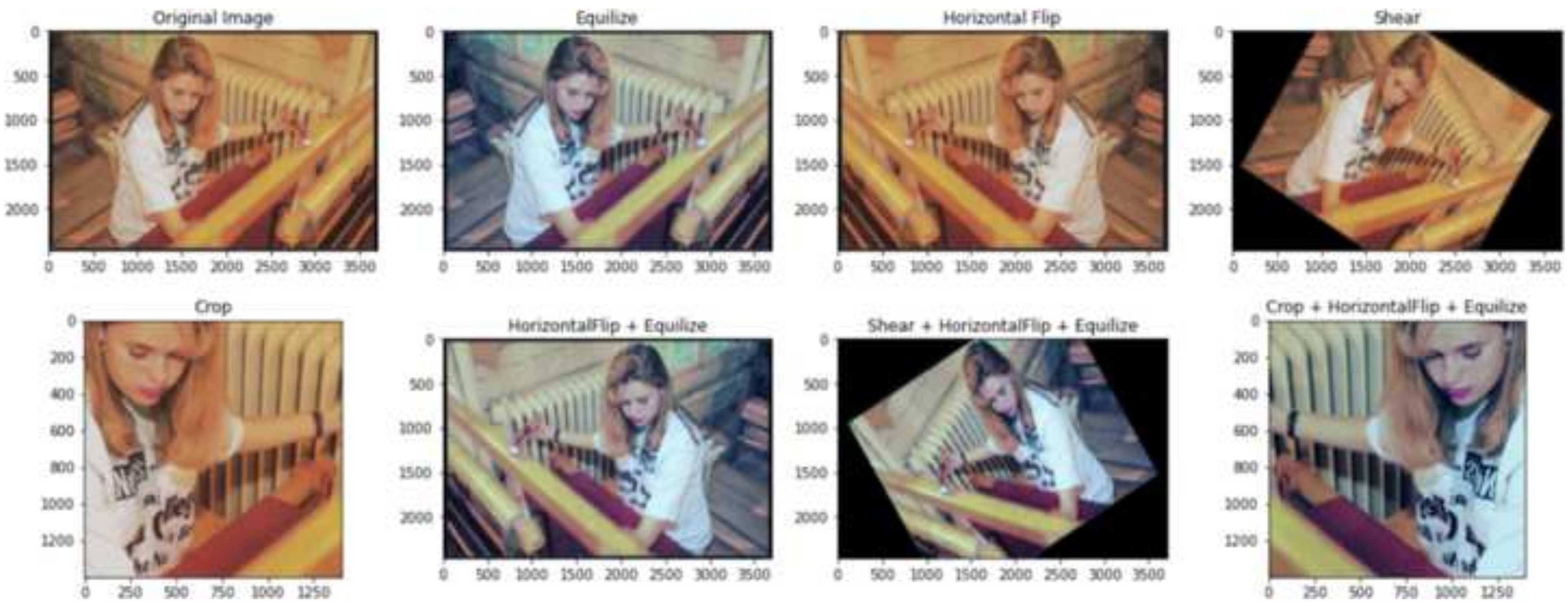


Figure 7: Overlapping male and female predictions

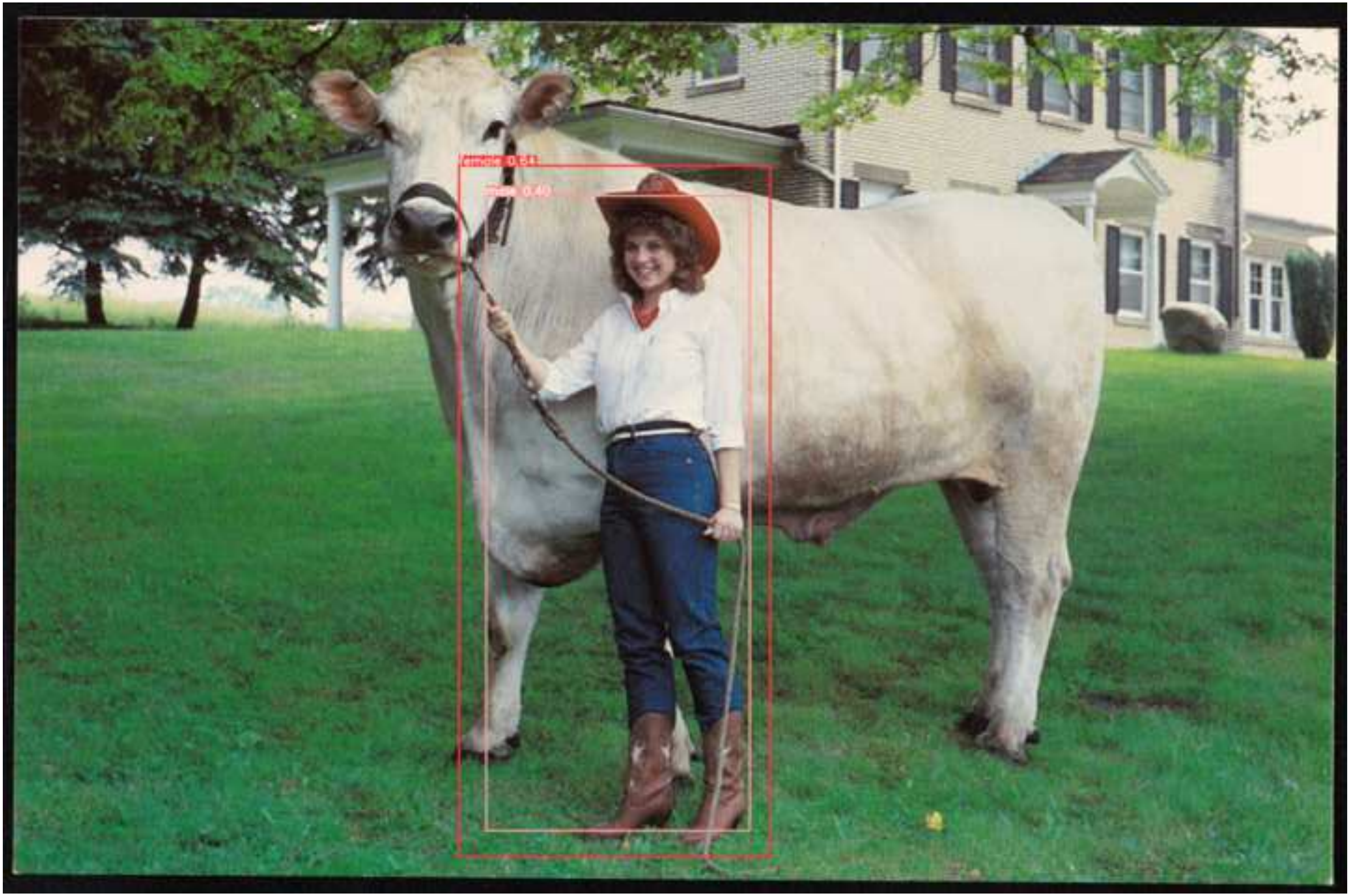


Figure 8: The intersection over the union (IoU) is calculated by dividing the area of intersection by the area of the union, producing a value between 0 (no overlap) and 1

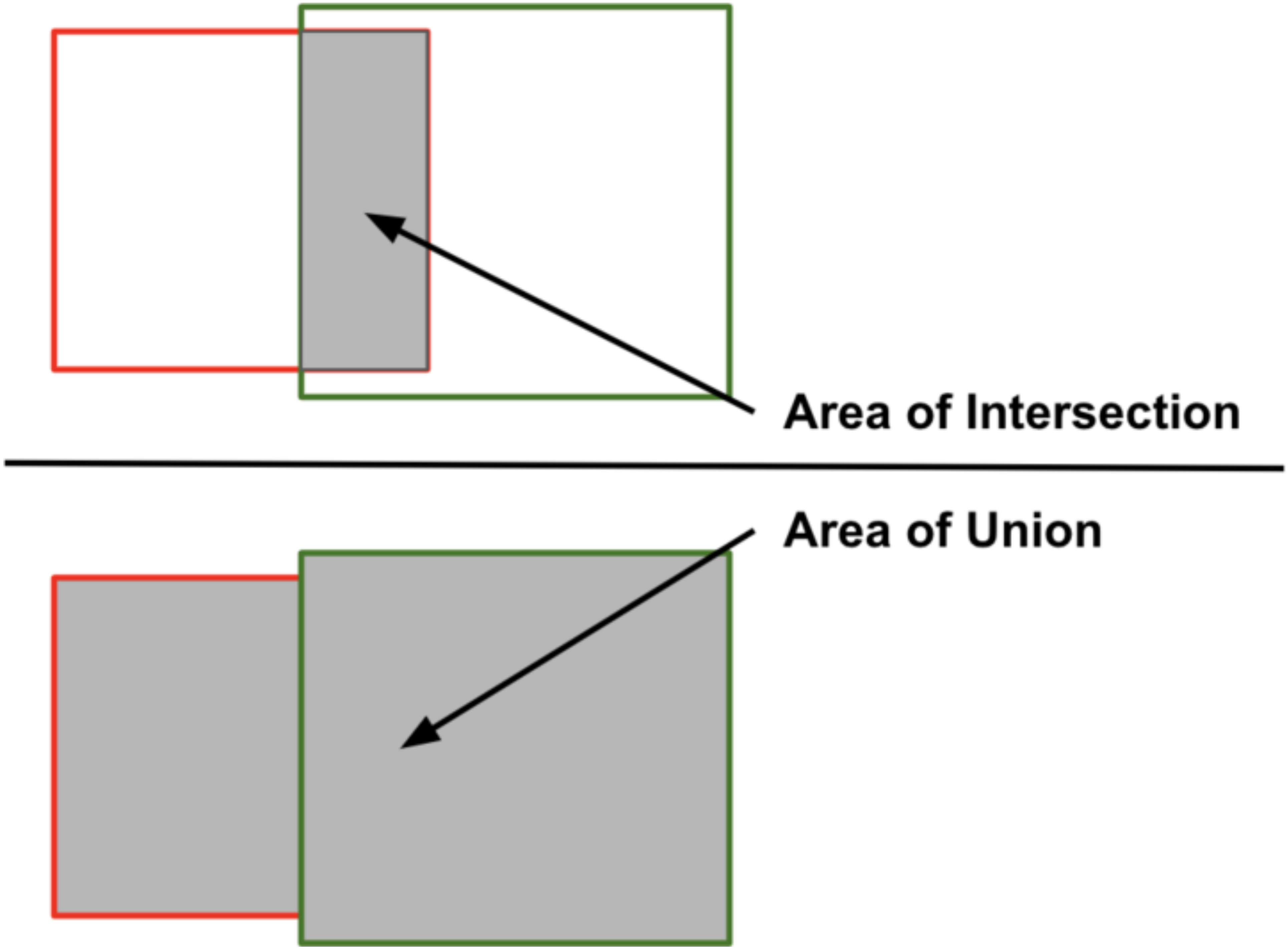


Figure 9: Example intersection over the union (IoU) values. The higher the value, the more the bounding boxes overlap.

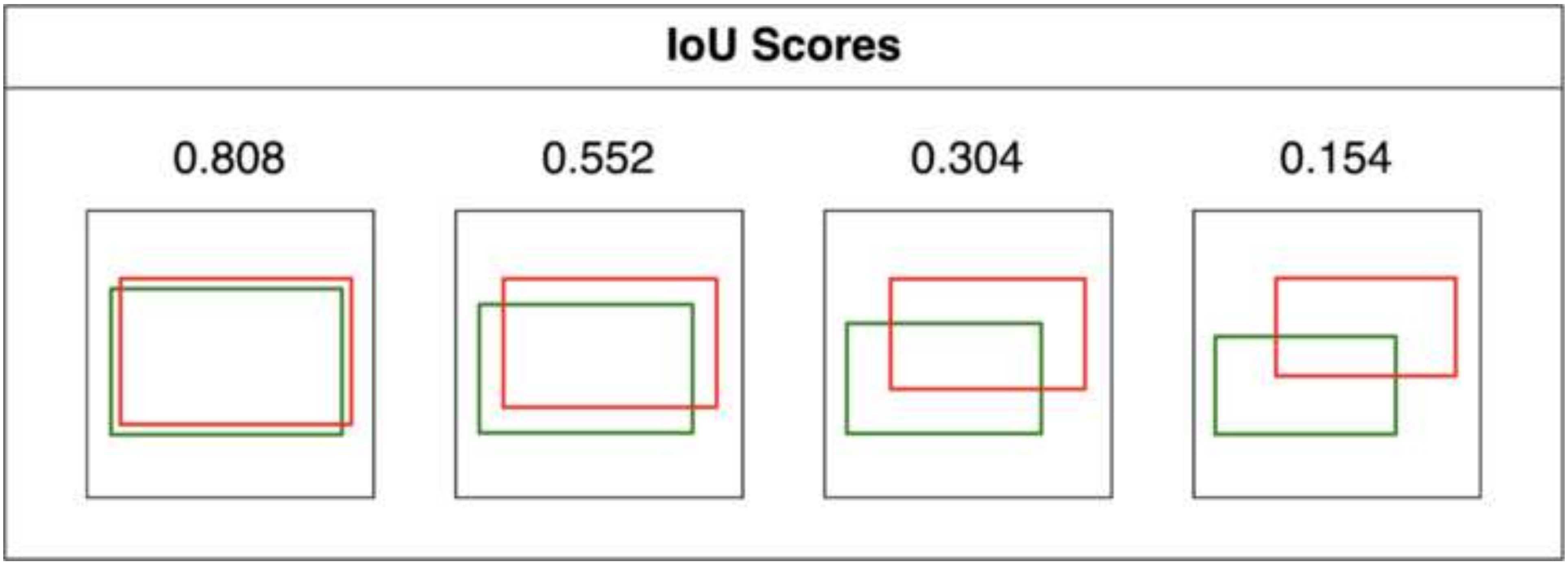


Figure 10: Breakdown of the 77 postcards removed from the testing set.

Click here to access/download;Figure;Figure 10.png

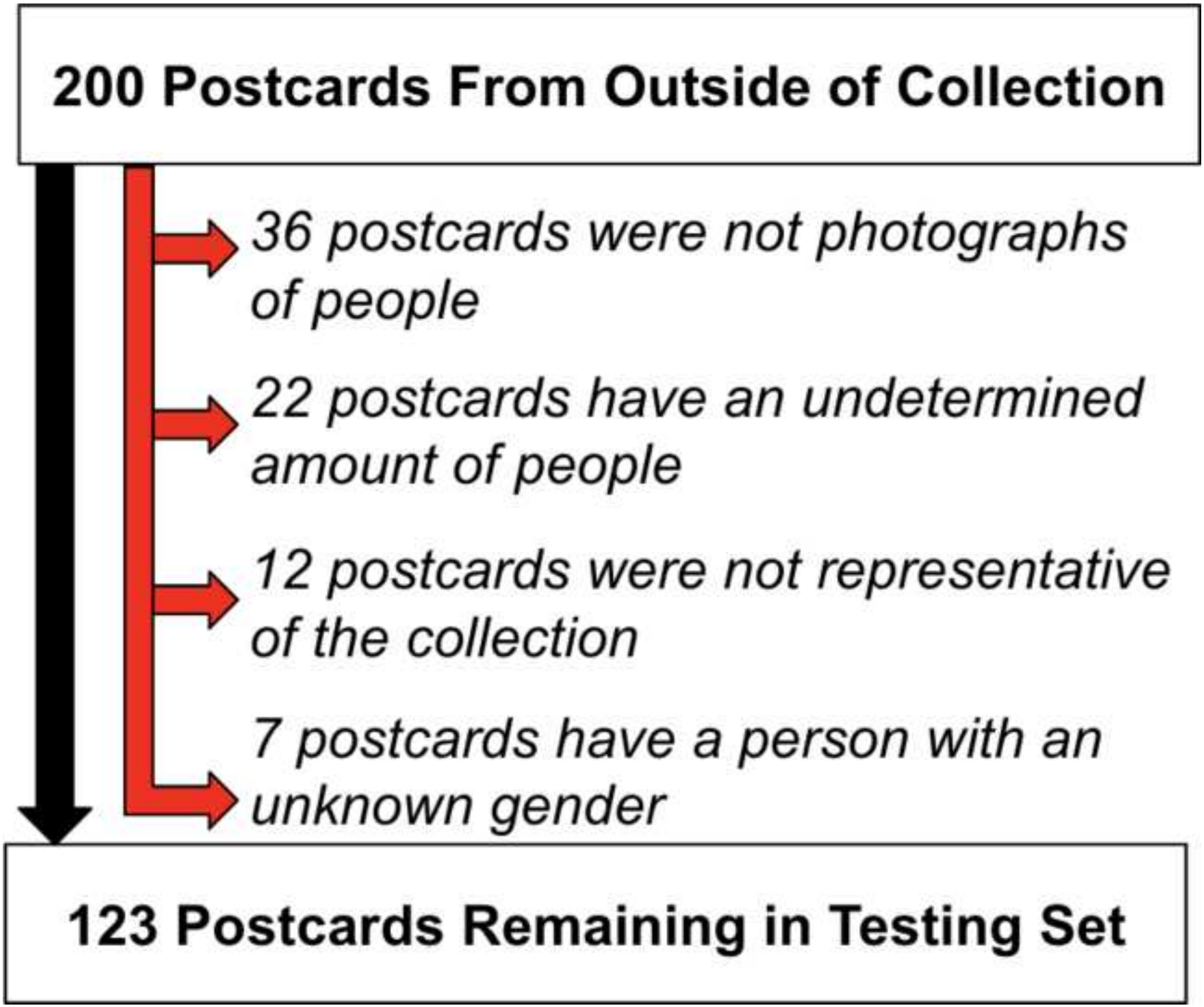
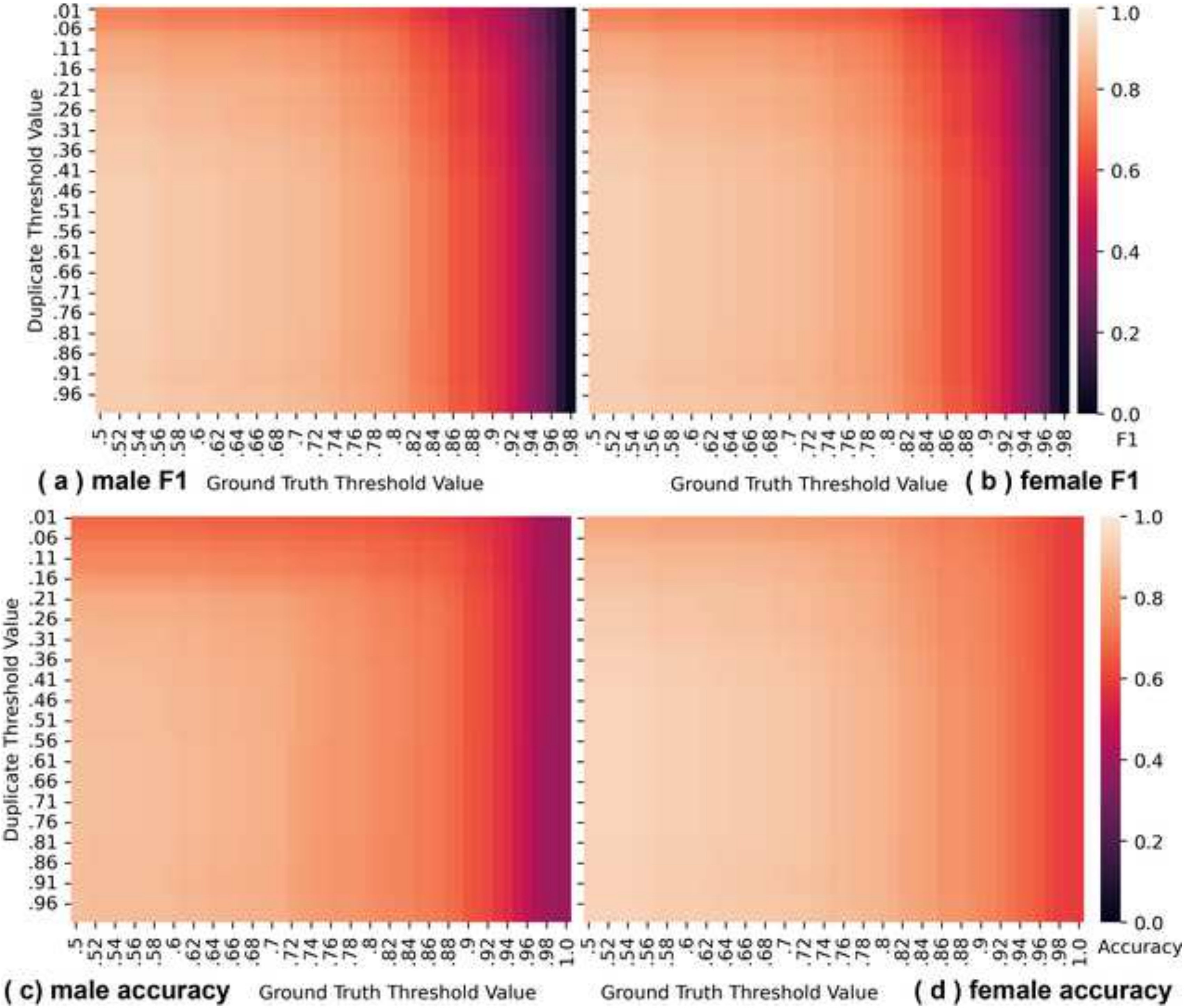


Figure 11: The effects of the IoU duplicate threshold and IoU ground-truth threshold on model F1 and accuracy for a model trained with its none of its layers frozen on the



List of Tables

Package	Version	Use
iterative-stratification (Sechidis et al., 2011)	0.1.7	Multi-label stratified 10-fold cross-validation
Pillow (Clark, 2015)	7.2.0	Reading and saving images
NumPy (Harris et al., 2020)	1.21.6	Reading images as array and calculations
Albumentations (Buslaev et al., 2020)	1.3.0	Data augmentation
pandas(Wes McKinney. 2010)	1.3.5	Reading, processing. modifying, and saving CSV files

Table 1: Python packages leveraged with their version and use.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Females in Training Set	1133	1133	1110	1113	1117	1109	1108	1110	1113	1096
Males in Training Set	1110	1080	1112	1124	1108	1123	1120	1121	1134	1110
Images in Training Set	881	890	893	888	887	883	886	893	885	888
Females in Validation Set	105	105	128	125	121	129	130	128	125	142
Males in Validation Set	128	158	126	114	130	115	118	117	104	128
Images in Validation Set	105	96	93	98	99	103	100	93	101	98

Table 2: The number of males, females, and images in the training and validation sets for each fold.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Females in Training Set	8665	8630	8488	8484	8579	8480	8528	8460	8505	8411
Males in Training Set	8495	8234	8462	8544	8437	8662	8589	8555	8645	8481
Images in Training Set	6845	6896	6920	6880	6872	6861	6878	6927	6863	6908

Table 3: The number of males, females, and images in the augmented training sets.

Model	Best IoU Duplicate Threshold	Best IoU Ground-Truth Threshold	Accuracy	Precision	Recall	F1
Augmented Unfrozen Fold 2	0.98	0.54	0.93939	0.91589	0.93333	0.92453
Unaugmented Unfrozen Fold 1	0.91	0.57	0.93562	0.94118	0.91429	0.92754
Unaugmented Unfrozen Fold 5	0.99	0.59	0.93227	0.97273	0.8843	0.92641
Augmented Backbone Fold 5	0.96	0.56	0.9243	0.93966	0.90083	0.91983
Augmented Unfrozen Fold 7	0.98	0.52	0.92339	0.944	0.90769	0.92549

Table 4: The best five models for female accuracy on their validation sets with their hyper-parameters optimized.

Model	Best IoU Duplicate Threshold	Best IoU Ground-Truth Threshold	Accuracy	Precision	Recall	F1
Unaugmented Backbone Fold 9	0.95	0.5	0.9214	0.91346	0.91346	0.91346
Augmented Unfrozen Fold 8	0.47	0.5	0.91736	0.96154	0.86207	0.90909
Augmented Unfrozen Fold 5	0.48	0.55	0.91667	0.936	0.9	0.91765
Augmented Backbone Fold 5	0.96	0.51	0.91633	0.97391	0.86154	0.91429
Augmented Unfrozen Fold 7	0.98	0.5	0.91532	0.95327	0.86441	0.90667

Table 5: The best five models for male accuracy on their validation sets with their hyper-parameters optimized.

Model	Best IoU Duplicate Threshold	Best IoU Ground-Truth Threshold	Accuracy	Precision	Recall	F1
Unaugmented Unfrozen Fold 1	0.91	0.57	0.93562	0.94118	0.91429	0.92754
Unaugmented Unfrozen Fold 5	0.99	0.58	0.93227	0.96429	0.89256	0.92704
Augmented Unfrozen Fold 7	0.98	0.52	0.92339	0.944	0.90769	0.92549
Augmented Unfrozen Fold 2	0.98	0.54	0.93939	0.91589	0.93333	0.92453
Augmented Unfrozen Fold 8	0.41	0.5	0.92149	0.928	0.92063	0.9243

Table 6: The best five models for female F1 on their validation sets with their hyper-parameters optimized.

Model	Best IoU Duplicate Threshold	Best IoU Ground-Truth Threshold	Accuracy	Precision	Recall	F1
Augmented Unfrozen Fold 5	0.48	0.55	0.91667	0.936	0.9	0.91765
Augmented Backbone Fold 5	0.81	0.51	0.91633	0.96581	0.86923	0.91498
Unaugmented Backbone Fold 9	0.95	0.5	0.9214	0.91346	0.91346	0.91346
Augmented Unfrozen Fold 8	0.47	0.5	0.91736	0.96154	0.86207	0.90909
Augmented Unfrozen Fold 7	0.98	0.5	0.91532	0.95327	0.86441	0.90667

Table 7: The best five models for male F1 on their validation sets with their hyper-parameters optimized.

Model	IoU Duplicate Threshold	IoU Ground-Truth Threshold	Accuracy	Precision	Recall	F1
Augmented Backbone Fold 4	0.89	0.5	0.69182	0.62162	0.39655	0.48421
Unaugmented Backbone Fold 2	1.0	0.53	0.6859	0.67742	0.35	0.46154
Unaugmented Backbone Fold 7	0.55	0.53	0.68027	0.64706	0.21154	0.31884
Augmented Unfrozen Fold 8	0.54	0.5	0.67133	0.57143	0.16327	0.25397
Augmented Backbone Fold 5	0.96	0.56	0.66474	0.73333	0.16923	0.275

Table 8: Best models for female accuracy on postcards from 1890-1919 in the testing set.

Model	IoU Duplicate Threshold	IoU Ground-Truth Threshold	Accuracy	Precision	Recall	F1
Augmented Backbone Fold 4	0.89	0.5	0.60265	0.69841	0.30345	0.42308
Unaugmented Backbone Fold 2	1.0	0.53	0.59603	0.73469	0.24828	0.37113
Unaugmented Backbone Fold 3	0.95	0.53	0.59091	0.66667	0.30872	0.42202
Augmented Backbone Fold 7	1.0	0.52	0.58571	0.71429	0.12	0.20548
Augmented Unfrozen Fold 7	0.98	0.52	0.58456	0.6	0.07759	0.1374

Table 9: Best models for female accuracy on all postcards in the testing set.

Model	IoU Duplicate Threshold	IoU Ground-Truth Threshold	Accuracy	Precision	Recall	F1
Unaugmented Backbone Fold 3	0.95	0.5	0.58824	0.94872	0.33036	0.49007
Augmented Backbone Fold 4	0.89	0.5	0.54088	0.91176	0.30693	0.45926
Unaugmented Backbone Fold 2	0.47	0.5	0.53846	0.77273	0.35417	0.48571
Unaugmented Unfrozen Fold 8	1.0	0.54	0.52439	0.83333	0.06098	0.11364
Augmented Backbone Fold 1	0.92	0.57	0.5102	0.33333	0.02857	0.05263

Table 10: Best models for male accuracy on postcards from 1890-1919 in the testing set.

Model	IoU Duplicate Threshold	IoU Ground-Truth Threshold	Accuracy	Precision	Recall	F1
Unaugmented Backbone Fold 3	0.95	0.5	0.63844	0.92727	0.32278	0.47887
Augmented Backbone Fold 4	0.89	0.5	0.61589	0.88679	0.29936	0.44762
Unaugmented Backbone Fold 2	0.47	0.5	0.57947	0.72059	0.3121	0.43556
Augmented Backbone Fold 10	0.55	0.51	0.52649	0.88889	0.10191	0.18286
Unaugmented Backbone Fold 5	0.94	0.59	0.51351	0.8	0.10256	0.18182

Table 11: Best models for male accuracy on all postcards in the testing set.