Pontifícia Universidade Católica
do Rio de Janeiro

Yan Martins Braz Gurevitz Cunha

Classifying Images with Unclear Patterns:

From Visual Features to Narrative Importance

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Informática.

Advisor: Prof. Sérgio Colcher

Rio de Janeiro
February 2023

Pontifícia Universidade Católica
do Rio de Janeiro

Yan Martins Braz Gurevitz Cunha

Classifying Images with Unclear Patterns:
From Visual Features to Narrative Importance

Dissertation presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Informática. Approved by the Examination Committee:

Prof. Sérgio Colcher
Advisor
Departamento de Informática – PUC-Rio

Prof. Alberto Barbosa Raposo
PUC-Rio

PhD. Antonio José Grandson Busson
BTG Pactual

Rio de Janeiro, February 16th, 2023

Yan Martins Braz Gurevitz Cunha

Graduated in computer science by PUC-RIO. Worked as intern software developer at Tecgraf and assistant researcher at Dexl Lab - LNCC. Current member of Telemidia lab

Bibliographic data

Martins Braz Gurevitz Cunha, Yan

Classifying Images with Unclear Patterns:From Visual Features to Narrative Importance / Yan Martins Braz Gurevitz Cunha; advisor: Sérgio Colcher. – 2023.

49 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2023.

Inclui bibliografia

1. Informática – Teses. 2. Classificação de Imagens; Aprendizado Profundo; Personagens de Anime; Narratologia. I. Colcher, Sérgio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

To my parents, for their support and encouragement, and for always giving me ideas and pushing me forward. To my friends who would stay with me on discord for many hours as I worked on this project and would listen to my complaints when things would go wrong. And to my closest friends in particular, whose emotional support would always help me keep going forward.

Acknowledgments

To my adviser Professor Sérgio Colcher for the support and guidance, always helping me find a way forward when the path was unclear. To my colleges at Telemidia, who, in many of our monthly meetings, helped with invaluable insight on how to achieve the goals of this project, especially to Ivan Jesus whose research in part inspired this work and who helped me to get it started. To everyone who took the survey, donating their time to help me find the answers to the questions posed here.

To CAPES, Telemidia and PUC-Rio, for the aids granted, without which this work could not have been accomplished.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Martins Braz Gurevitz Cunha, Yan; Colcher, Sérgio (Advisor). **Classifying Images with Unclear Patterns: From Visual Features to Narrative Importance**. Rio de Janeiro, 2023. 49p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The field of image classification has been heavily explored for years, especially with the big advancements in deep neural networks seen in the last decade. However, most of the focus has been dedicated to cases with significant inter-class differences and minor intra-class differences. In this work we explore how well convolutional networks deal with cases with small inter-class differences and whose classification carries a degree of subjectivity, making non-obvious the relationship between visual features and classification and differentiating it from the traditional field of fine-grained classification. To do that, we approach a specific instance of this problem: Determining a character's narrative importance based solely on its image. We have evaluate the performance of different CNN models in our task, using a dataset we created for it, and we have analysed which patterns were found when it comes to the relationship between visual features and classification. We show that, for our specific task, CNNs are able to exceed human performance in pure accuracy and, more interestingly, mirror many of the patterns humans show when judging characters, even if some of those patterns are inaccurate. This means that this kind of model may be able to serve as a good surrogate for human evaluators when designing characters.

Keywords

Image Classification; Deep Learning; Anime Characters; Narratology.

Resumo

Martins Braz Gurevitz Cunha, Yan; Colcher, Sérgio. **Classificando Imagens com Padrões Incertos: das Propriedades VIsuais à Importância da Narrativa**. Rio de Janeiro, 2023. 49p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O campo de classificação de imagens tem sido bastante explorado há anos, em especial com o grande avanço de redes neurais da última década. No entanto, grande parte do foco tem sido dedicado a casos com grandes diferenças inter-classe e pequenas diferenças intra-classe. Neste trabalho exploramos o quão bem redes convolucionais lidam com casos com pequenas diferenças inter-classe e cujas classificações carregam um grau de subjetividade, torando não óbvia a relação entre features visuais e classifcação e diferenciando isso do campo tradicional de fine-grained classificaiton. Para isso, abordamos um caso específico deste problema: Determinar a importância narrativa de um personagem a partir somente de sua imagem. Avaliamos a performance de CNNs em nossa tarefa, usando um dataset que criamos para ela, e analisamos que padrões conseguimos encontrar no quis diz respeito da relação entre features visuais e classifcação. Mostramos que, especificamente para a tarefa que estudamos, CNNs conseguem superar a performance humana em termos de acurácia e, além disso, refletem vários dos padrões apresentados por humanos quando julgam personagens, até mesmos alguns padrões que não refletem a realidade. Isto significa que esse tipo de modelo pode ser um possível substiuto para avaliadores humanos para propósitos de character design.

Palavras-chave

Classificação de Imagens; Aprendizado Profundo; Personagens de Anime; Narratologia.

Table of contents

List of figures

List of tables

List of codes

List of Abreviations

CNN – Convolutional Neural Network

1
Introduction

Classifying images based on their visual features has been a fixture of Computer Vision for many decades, with various approaches leading to significant performance gains over the years. However, most of the effort on the field has been devoted to cases with significant inter-class variations and minor intra-class variations, making it easier to find clear patterns to differentiate between the classes (1), as exemplified by the prominence of the Imagenet dataset (2) as a benchmark for different techniques developed over the last decade. In this work, we seek to explore how image classification models based on Convolutional Neural Networks (3) perform when trying to classify images with patterns that are unclear to humans, as well as considering if this classification can be used to better understand these patterns in both model and human interpretation.

Classifying images with unclear and sometimes implicit patterns has many applications, especially in tasks related to analyzing human features and behavior, as some correlations can be hard to determine (4). For this work, we chose to focus on a specific example of this kind of task: based solely on an image of a character, classify it according to its narrative importance.

1.1
Narrative importance classification

A story can be defined as the actions of characters; in that sense, a character's actions determine what they are to the narrative. For that reason, it has been common to classify characters according to their role in the overall plot, as seen in (5, 6, 7), as it helps both the author and audience know what type of actions to expect from that role. It has also been pointed out how visual appearance impacts how we perceive other people (4) and, extrapolating to the context of narrative media, the correlations between how a character looks and the expectations it creates in the audience (8, 9) and how it relates to their narrative role (10).

However, as Rogers et al. (10) mention, the traits one expects from a given role can vary from person to person. At the same time, the authors found a lot more consensus once the task changed to determining a character's position based on their design. This indicates that the strong correlations between visual and narrative elements are complex for humans to define explicitly, making it an excellent example of classification with unclear patterns.

In this work, we focus on one version of this problem: based purely on an image of a character, trying to determine whether it's one of the story's main characters or just part of the supporting cast. We evaluate how models based on Convolutional Neural Networks (CNNs) (3) perform on this task while comparing it to human performance via a survey. We also try to find patterns in the correlations between character design and narrative importance. We also consider what these patterns tell us about the specific task and the general classification field with unclear patterns.

We chose to limit the scope of our research to anime characters due to the relative visual simplicity, and exaggerated traits (11) when compared to more realistic images, as well as helping to maintain visual consistency thanks to being limited to one type of media.

2
Related Work

This chapter will cover previous works related to image classification with unclear patterns in general and the specific example of narrative importance classification.

2.1
CNNs and Image Classification

Over the past decade, we've seen the resurgence of deep neural networks, with CNNs (3) proving particularly effective in computer vision tasks due to their ability to extract image features without needing every layer to be fully connected. The initial success of Krizhevsky et al. (3). sparked a wave of research in this type of architecture, leading to consistent advances over the years. The second half of the decade saw the emergence of residual models (12) that would dominate many tasks like Super-Resolution (13) and, more importantly for this work, image classification. Residual networks could learn much more complex functions and have much greater depth without suffering as much from the vanishing gradient problem, thanks to their skip connections.

CNNs, more specifically ResNets, have also been used to classify anime characters according to their emotions (14). While this use doesn't fit the type of task we studied, it is tangentially related to the specific task of classifying anime characters according to their narrative role and shows that this kind of architecture can be used in tasks relating to the medium of anime.

2.2
The case of narrative importance classification

While there were a few other works correlating character design and narrative role in narratology, CNNs (3) have seen a lot of study in the past decade. There have been studies on character generation (15), especially after the growth of GANs (16). We could not find any research combining the ideas of character design and narrative role in the context of deep learning-based classification and data study, so we first had to approach each topic independently.

2.2.1 Narratology

While in this work we chose to analyze characters based on their narrative prominence, it's important to note the relationship between specific narrative roles and their narrative weight (6), as a hero is more likely to be the main character than a trickster. These correlations may help us understand the results we find later. For this purpose, we followed the definition of narrative roles as a set of functions that can be performed by attributes that can belong to a narrative entity (usually, but not limited to, a character) (17). Each action performed by an entity can be seen as a function of the narrative role that entity is playing, thus revealing the dynamic nature of narrative roles, as one entity can play many different roles throughout the story (6).

While we define roles as sets of narrative functions and attributes, it is essential to make clear that a given function or attribute can belong to multiple roles and even change its nature depending on which role performs them (for example, a hero killing another character has an entirely different meaning than if a villain were to do the same). At the same time, the same group of functions and attributes that define a role can change significantly, as exemplified with visual traits in (10), and which of those elements is strictly required for an entity to fit a role is up for even further debate. All of that points to the fact that what makes audiences identify characters as a given role is linked to implicit expectations present in each audience member, thus making the task of learning this correlation even harder.

As for how many different roles exist and how we define each one, there have been many attempts to create typologies of narrative roles, with some of the most notable being Campbell's archetypes (5), their refinement by Vogler

(6), and Propp's dramatis personae (7). All of these share a few roles, like the hero and the villain (or shadow).

2.2.2 Character Design

In the field of character design, we have seen attempts to correlate visual and narrative elements in different types of media, such as Rogers et al.'s work (10) on video game Non-Player Characters which helped find correlations between certain visual traits and their expected narrative role; Hoffner and Cantor's studies (8, 9) on how factors such as appearance impact the expectations towards television characters, helping to solidify the idea that there is an essential correlation between narrative and visual elements. More closely related to the focus of this work, Minghua Liu and Ping Wang's analysis (11) of visual design in animation, particularly regarding how they establish observable traits expected from main characters in Japanese animation.

3
The Data

To study how CNNs perform on the narrative importance task, we had to create our own dataset from images and labels found online.

Finding anime characters' images online is easy enough, with many sites containing images of over 100,000 characters. To that end, we used the pictures from MAL (MyAnimeList.com) because the platform also classifies every character between Main and Supporting according to their narrative prominence, thus providing our target labels. Using that, we created an initial dataset of over 31000 characters, each represented by a single image.

We used CLIP (18) to filter out any characters not classified as human. This was done to remove from the data characters with mostly animalistic, fantastical, or mechanical appearances, as such elements could affect the impression they create on audiences and make it harder to use the ideas of person perception (4) in the analysis.

3.1
Distribution of Visual Features

The final dataset consisted of 18,550 images, with 53.24% of them belonging to main characters and 46.76% to supporting characters. For the CNN-based classification, the data was divided between training (70%), validation (20%), and testing (10%) groups, while maintaining the exact proportions of both classes in each group. The latter subset would also be used in the human survey to compare human and model performance better.

Following Minghua Liu and Ping Wang's description of Main characters in Japanese Animation (11), we decided to focus on the visual features most exaggerated by the medium, those being hair, eyes, apparent gender, and apparent age.

3.1.1
Apparent Gender

While the actual gender a character identifies as in the narrative may differ from their appearance, for this task, we limited ourselves to their apparent gender as classified by CLIP (18) since we are working with only their character design without external narrative context.

This left us with 45.20% of the characters being classified as male and 54.80% as female. The relative balance was expected due to the varied nature of demographics the medium of anime appeals to and the importance each gender has in stories that appeal to the opposite. However, the prominence and role a character of a specific gender plays may change depending on the narrative genre and target demographic.

3.1.2
Apparent Age

Similarly to gender, a character's actual age may differ from their appearance, but again we focused on their apparent age since that is all the model and the survey takers had access to. Here the division was a lot more imbalanced than in the case of gender, with 87.43% of characters being classified as young and 12.57% as old. However, this imbalance is not surprising given that audiences tend to associate old age with few narrative roles beyond that of the mentor (10), which can even be seen in Campbell's original name for the archetype as the Wise Old Man or Woman (5).

3.1.3
Eyes

Exaggerated eyes are an important element in both American and Japanese animation, with the latter showing more intensity and frequency in how it uses the concept (19). The importance of eyes as an indication of a character's mental, emotional, or narrative state in the medium of anime comes as a consequence of the role they play in Japanese culture as a representation of one's emotions (20). Therefore it's not surprising that their size would be an essential factor in defining the appearance of a main character (11).

The fact that the use of this feature in anime originated in Shojo (teenage girls) manga (21), alongside the importance (11) gives to it in the specific context of female main characters, acts as a strong indication that the gender of the character may heavily influence this feature.

We used Illustration2Vec's tag extractor (22) to get a character's eye size by using the model's likelihood of them having closed eyes, with the higher values meaning smaller eyes. Main characters have, on average, half the likelihood of having closed eyes compared to supporting characters.

3.1.4
Hair

As Minghua Liu and Ping, Wang (11) note, the importance of hair in anime character design, especially in the case of female main characters. As color plays an important role in conveying narrative meaning in the medium (23), we measured how factors such as hair length and color impact the classification between leading and supporting characters.

We again used Illustration2Vec (22) to classify the characters according to hair length and CLIP (18) to do the same for hair color. In the length measurement, 64.78% of the characters were classified as having short hair, and 35.22% as having long hair. Still, this distribution is heavily impacted by apparent gender, as over 55% of female characters had long hair. How this feature affects the narrative classification for male and female characters will be seen when we analyze the results of both the model's classification and the human survey.

The distribution of hair colors can be seen in figure 4. The fact that over 60% of characters feature usual hair colors such as black, brown, blonde, and grey shows an interesting contrast to anime's fame of having characters with unusual and intense hair colors.

4
Human Survey

Due to the unclear nature of the patterns present in the data, alongside the lack of conclusive research on the topic, we decided it was important to measure the human performance on the task of classifying a character between main and supporting based solely on their image. The intention was for this to act as baseline against which we could compare our models.

Going into the survey, our expectations were mixed. On one hand it has been mentioned that audiences are better at identifying a character's role than they are at determining which visual features indicate each role (10), but at the same time the correlations in the data between visual and narrative features were unclear, making us unsure of which patterns survey takers would follow. Therefore, besides their quantitative performance, we were also interested in seeing which biases the survey takers would carry with them.

The survey as conducted online and with people familiar with the medium of anime. Each participant was shown one character at a time and had to classify each one according to whether they thought it was a main or supporting character, while also being given the option to say that they already know the character, to make sure that the valid classifications were based sole on the image shown. The characters shown to survey takers were pulled from the subset of the data dedicated to the testing of the CNN model and were shown in a different random order to each taker. Each participant could classify as many characters as they wanted and could save their progress and return to the task whenever they wanted, as to avoid the impact of fatigue on their performance.

In total we had 21 participants totalling over 6000 individual classifications, possibly giving us a good indication of which factors people familiar with the medium consider when trying to first identify the importance of a character.

Since every character got more than one classification, we analyzed the data by considering the majority answer for each character.

4.1 Overall Human Performance

As for how the test takers performed when compared with the ground truth, the total accuracy was 58.82%, showing that people are somewhat able to determine how important a character is at first glance, but also that many of their biases may be hindering their judgement, be it because such preconceptions are simply unfounded or because they're purposefully subverted by authors and designers.

An interesting pattern that quickly emerged is that, while over 53% of characters were classified as main in the ground truth, when considering the majority answers for each character, humans classified only 44.55% of them as main and 55.45% as supporting, indicating that humans might have more demanding expectations of what main characters should look like, which would be reinforced by the somewhat strict definition (11) gives to the role.

Using other metrics to evaluate this discrepancy, we saw that the precision for supporting characters was considerably lower than that of main characters (54.44% to 66.04%), indicating a higher quantity of false positives. Meanwhile, the recall for supporting characters was much higher (72.54% to 46.80%), contributing to the image of low coverage in the main class. However, it's important to reiterate that, while these numbers paint an image of low performance on the part of the survey takers, they are consistent with the strict definition given to the class in general when it comes to its visual design (11), as well as with the specific expectations of each visual elements, as we'll see when analysing the results.

Table 4.1: Metrics for the different classes

	Precision	Recall	F-Score
Main	0.6604	0.4680	0.5478
Supporting	0.5444	0.7254	0.6220

5
CNN-based Classification

To measure how CNNs perform on the task and which patterns they show in their correlation of visual features to narrative importance, we trained different models using the data subsets detailed in 3, keeping the same subset used for the human survey as the testing subset.

The models we chose mainly focused on different ResNets (12) and the SpinalNet (24). We alternated between training models from scratch or fine-tuning pre-trained models, while at the same time testing different parameters like optimizers, learning rates, and different techniques for data augmentation. We found that larger models we quickly overfit and had to run low learning rates to avoid this problem. This issue can be caused by the indirect relationship between visual features and labels and could turn up again on other tasks of this kind.

Initially, our best results came from a ResNet50 pre-trained on the imagenet dataset (2) and fine-tuned with the rmsprop optimizer, achieving 64.56% accuracy on the testing subset, outperforming the humans surveyed by close to 6%, so we'll use this model to compare the patterns in visual features.

In contrast to how humans only classified 44.55% of the characters as main, the model stuck a lot closer to the 53.24% proportion shown in the ground truth and classified 52.64% of characters as main and 47.36% as supporting. This can also be seen when considering that the precision and recall between the two classes were a lot closer for the model, with the former metric being 61.95% for supporting characters and 66.90% for main characters and the latter being 62.74% and 66.16% for supporting and main characters respectively. This helps indicate that the model may be considerably less impacted by the biases that hinder human performance, but we'll have a clearer picture of this when comparing their results in each visual feature.

Table 5.1: Metrics for the different classes

	Precision	Recall	F-Score
Main	0.6690	0.6616	0.6653
Supporting	0.6195	0.6274	0.6234

We also decided to compare the loss achieved by the model with that of humans, for a more detailed performance comparison. To calculate the cross entropy of the survey results we considered the likelihood of a character belonging to a class being the number of times he was labelled as a member of it divided by the total number of times the character was classified. Using that method the loss of the survey was 3.84 while the loss of the model was 0.63, showing a much larger difference in performance.

5.1 The Use of Attention

In recent years the attention mechanism has been the focus of a lot of study and helped reduced computational cost and improve performance in many image classification tasks (25). Considering that, we added attention layers to the end of our pre-trained ResNet50, similarly to what has been done to achieve significant gains in other classification tasks (25). On top of improving performance, we were also interested in seeing what the attention mechanism would focus.

The initial results were slightly better, with a test accuracy of 65.26%, but the gains were a lot lower than expected.

The explanations for the disappointing results lie, in part, in what the attention mechanism ended up focusing. Most of the focus was given to the characters' eyes, which can be significant indicators of their importance, as we'll explore in chapter 7, but on their own aren't enough to achieve great performance.

6
MLP Classification

In order to compare how well our CNN's feature extraction performs against a scenario where we manually choose which features to consider, we trained MLP models on a categorical version of the dataset. We used the visual features detailed in chapter 3 and alternated which ones were included each time, gradually removing those that were seen to have a weaker correlation with the label. We used the following light model, to allow for quick training while serving as baseline for the task.

Code 1: MLP Model

```
model = Sequential()
model.add(Dense(10, activation='relu', kernel_initializer='he_normal', input_shape=(nFeatures,)))
model.add(Dense(8, activation='relu', kernel_initializer='he_normal'))
model.add(Dense(1, activation='sigmoid'))
```

For the training we used the Adam optimizer (**??**) and settled on learning rate of 4e-3. We trained for 1000 epochs on each permutation of features and results were the following.

Table 6.1: Results for MLP models

	All Features	Age, Hair color and Length, and Eye Size	Age, Hair color, and Eye Size
Accuracy	0.620	0.615	0.612
Loss	0.650	0.655	0.652

The results were, in terms of raw accuracy, inferior to our CNN model, showing that its trained feature extraction is superior when it comes to this task. There is, however, more to be seen when analyzing how the different models and the survey takers relate when it comes to the correlations between visual features and label, so in the next chapter we'll explore the pattern found in this regard.

7
Results

While we've already established that, in terms of accuracy, both the ResNet model and the MLP models outperform the humans surveyed, we want to see in more detail how the different methods of classification compare, which patterns in terms of visual features they follow, and how it all relates to the patterns present in the data.

7.1
Different answers between humans and the model

To analyze the patterns in visual features and how they relate to narrative importance, it's important that we first know the context of how close were our classifiers in their answers, and how they relate to the ground truth.

In most cases, the survey takers and the ResNet model were in agreement, with both groups giving the same classification to a character in 66.66% of cases, out of which 70.22% were correct. When the groups differed in their responses they were fairly closely matched, with the model being correct in 53.23% of cases and the humans in 46.77%. The fact that the percentage of agreement between the two groups was higher than the accuracy of the model reveals an interesting fact, that the model was slightly better at mirroring human pattern perception than it was at learning the patterns actually present in the data.

However, when we look at the same numbers for each class we start to see more clear patterns. By firstly looking at the supporting characters, we see that the rate of agreement was slightly higher, at 68.74%, out of which both groups were right in 73.15% of cases. But the most important value is when the two groups gave different answers, with the model being correct in only 39.85% of those cases, while humans were in 60.15%. This reflects the fact that, as previously mentioned, humans were more likely to classify characters as supporting, when compared to both the model and the ground truth, making them less to mistakenly classify a supporting character as main. At the same time, when it comes to the main characters, they agreed in only 64.84% of cases, and out of the ones they disagreed the model was right in 63.68%, meaning humans were a lot more likely to label a main character as supporting. Now it's important to find which visual factors were the most relevant in determining those differences in classification.

The MLP models painted a completely different picture, however. All three agreed with the survey takers on just around 40% of cases. This shows that, while in terms of accuracy they might have been close to the ResNet, they are unsuitable as representatives of human performance. Therefore, when analyzing the visual patterns found, we'll only consider the ResNet model.

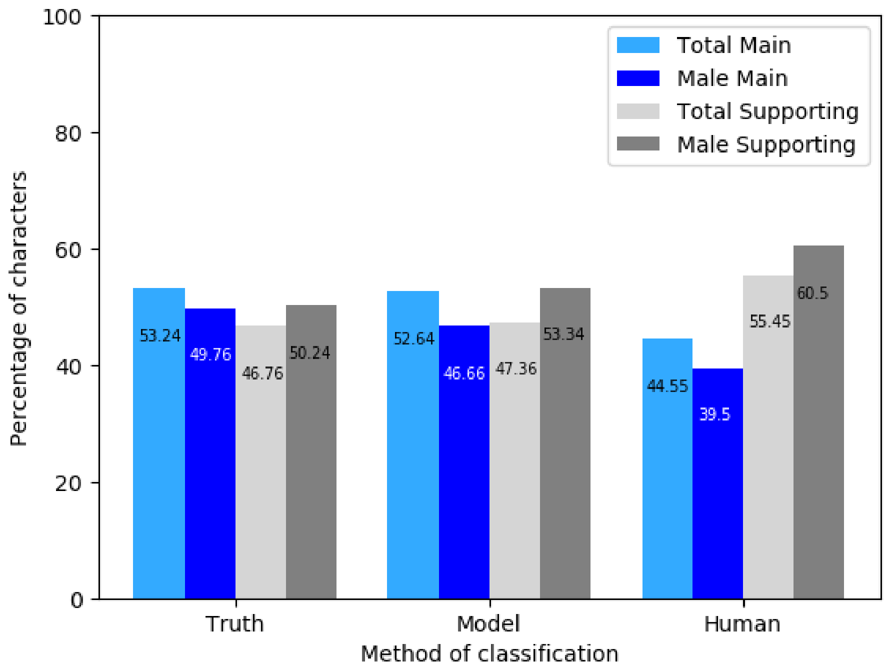
7.2 Apparent Gender

When considering the distribution of the classes according to this feature, the male characters 49.76% were main characters and 50.24% supporting, which initially seems to indicate a balance between the groups, but when we consider that 53.24% of all characters were main characters, being classified as male represents a 6.54% lower chance of being a main character. At the same time, 56.10% of female characters were main characters, a 5.37% increase, which shows that a character being female makes it more likely for it to be a main character. However, while the difference is relevant, its relatively low magnitude is a reflection of the diversity in demographics the medium of anime tends to target (26), with different series targeting people of different gender and age groups.

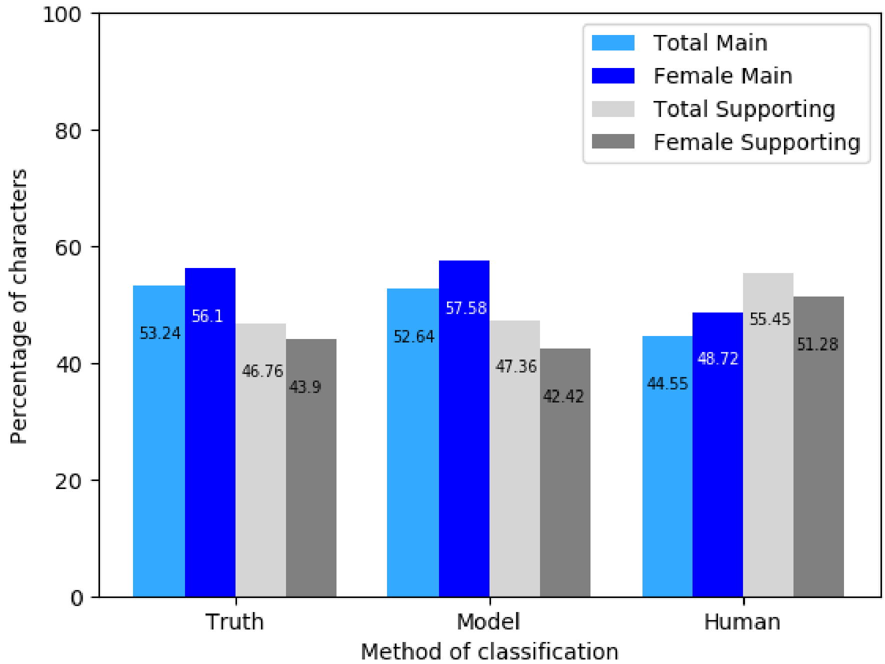
The model exacerbated the trend, with male characters being 11.36% less likely to be classified as main and female ones 9.38% more likely to receive the same label. And humans showed similar results, with being male resulting in 11.23% lower likelihood of being a main character and being female a 9.36%

higher likelihood.

This shows that both humans and the model consider a character's apparent gender to be more indicative of their importance than it actually is, with the model giving the most importance to the feature. It will be important to keep this factor in mind when analyzing how the other features are impacted by gender, especially considering that expectations of how a character should look are heavily influenced by it (11).



(a) Male class distribution, when compared to all characters



(b) Female class distribution, when compared to all characters

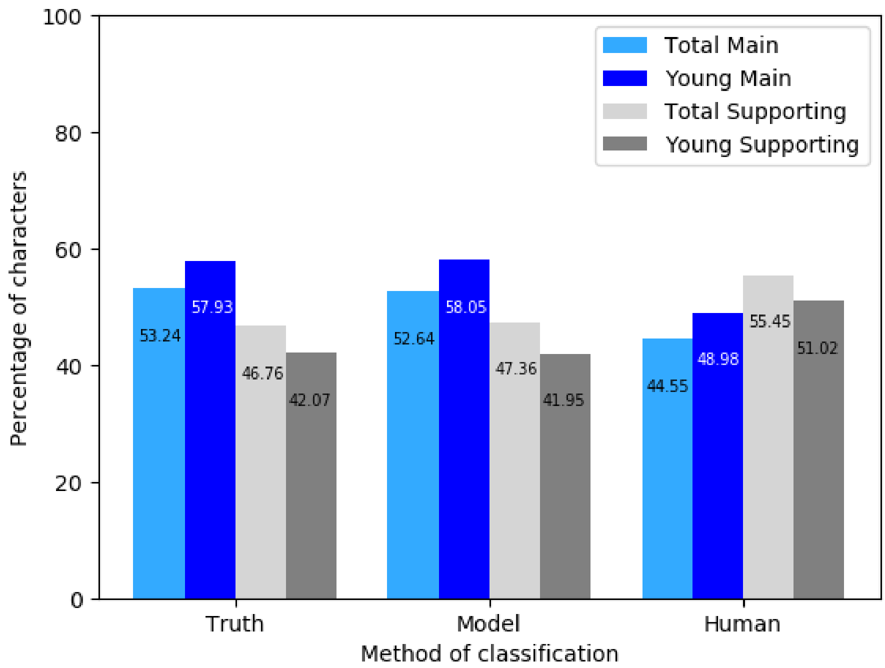
Figure 7.1: Graphs showing how apparent gender impacts class distribution.

7.3
Apparent Age

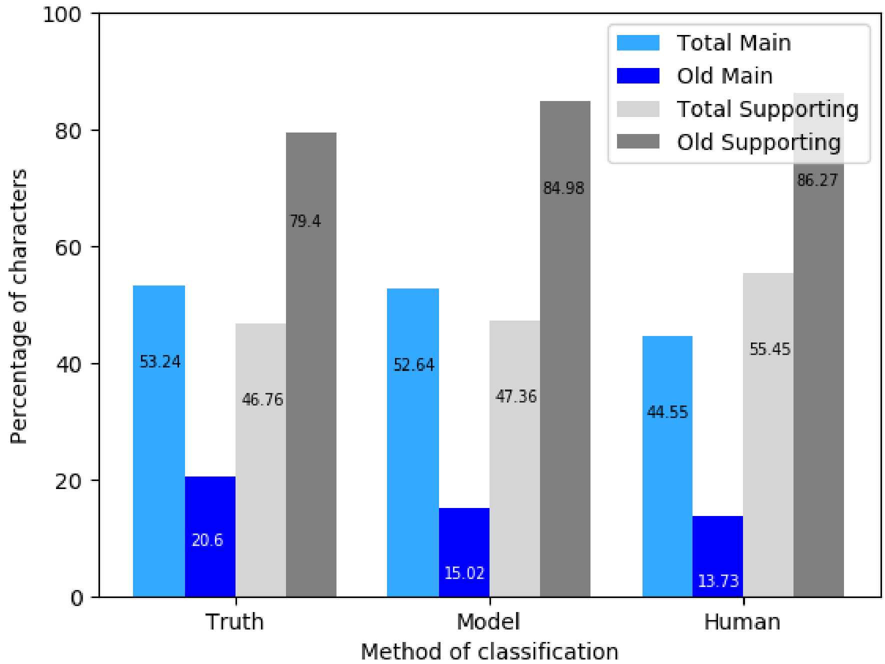
When analyzing how apparent age impacts the classification of a character, it's important to remember that, as mentioned in section 3.1.2, a large majority (87.43%) of characters were considered young, so the distribution of classes in this age group will most likely be a lot closer to the overall distribution than when considering old characters.

Out of the old characters, only 20.6% were main characters, representing a 61.30% lower chance of belonging to the class. This distribution may seem extreme at first, but makes sense when we consider that old age is usually associated with wisdom and the role of mentor (10) and that these types of characters tend to take on supportive roles (5, 6), helping the protagonist and their close allies. Added to that the fact that such characteristics usually mean that a character has already gone through an arc or journey of their own, making it easier for authors to avoid showing their development in favor of focusing more on the protagonist(s), by making it feel like these characters don't need any further development (27). Another important factor to consider when looking at this data specifically for the medium of anime is that none of the main core target demographics (Shonen, Shojo, Seinen and Josei) are aimed at older people (26), with the first two being mostly focused on teen boys and girls respectively and the latter two on young men and women respectively. And as the main characters are the ones the audience is supposed to more closely identify and relate with (6), it's natural that they are more likely to reflect their target demographic.

Both the model and the humans surveyed showed more intense versions of this trend, with the latter considering a character 69.18% less likely to be a main character if they were old and the former again showing the most extreme results as it considered an old character to be 71.46% less likely to be a main character when compared to the average. While these results are more pronounced than the actual data, they help illustrate how important a character's age is when determining how important they'll be and they give us our first concrete answer when it comes to what anime main characters tend to look like, they are overwhelmingly young. These numbers didn't change much when considering each gender in isolation, which matches how being physically attractive is usually seen as an indication of being good and important (10, 11) for both male and female main characters.



(a) Young class distribution, when compared to all characters



(b) Old class distribution, when compared to all characters

Figure 7.2: Graphs showing how apparent age impacts class distribution.

7.4 Eyes

The expectations for this feature were that main characters would be more likely to have big eyes, since these are used to communicate a character's emotions (20) and the protagonists are the ones whose emotions and mental state tend to receive the most attention (6). In addition, big eyes can be correlated with being good or innocent (20), traits that are associated with the role of hero (10).

When looking at the results, our expectations were initially proven correct, with supporting characters having a 122.4% higher likelihood of featuring small eyes. The model once again showed an exaggerated version of the pattern, with supporting characters being 165.58% more likely to have small eyes. However, this time humans represented the most drastic results, with supporting characters 181.91% more likely to have small eyes. This indicates that, while eyes are an important indicator of a character's importance, the importance of this feature might be overestimated by the audience which may expect main characters to look good and innocent.

When looking at each gender in isolation, female characters showed lower differences between main and supporting characters, with the latter group having a higher likelihood of featuring small eyes of 29.72%, 136.8%, and 158.69%, according to the ground truth, the model, and the human survey, respectively. This indicates that eye size isn't as much of a distinguishing factor when it comes to female characters, but the fact that humans and the model didn't follow the intensity of the real drop in difference, shows that there's still an expectation that having big eyes is a strong indicator that a female character is a protagonist, when in reality female characters were shown to be more likely than their male counterparts to have big eyes, with even the average of female supporting characters outperforming the same metric for male main characters. Therefore, having big eyes as a female character is not exclusive to main characters and the feature is widely present in both groups, which means that saying that female protagonists tend to have big eyes (11) isn't wrong, but it's also not very helpful in finding a pattern for the role since female characters, in general, tend to have bigger eyes.

When looking at only male characters, those labeled as supporting showed a 182.9% higher chance of having small eyes, revealing that, differently from what the author suggests in (11), the feature is more relevant in determining a character's importance when they are male instead of female since having big eyes is a rarer trait in the former while being quite common in the latter. At the same time, its rarity in male characters means that there are enough male main characters with both small and large eyes, making it so that having big eyes makes one a lot more likely to be a main character, but having small eyes doesn't carry the same effect in the opposite direction.

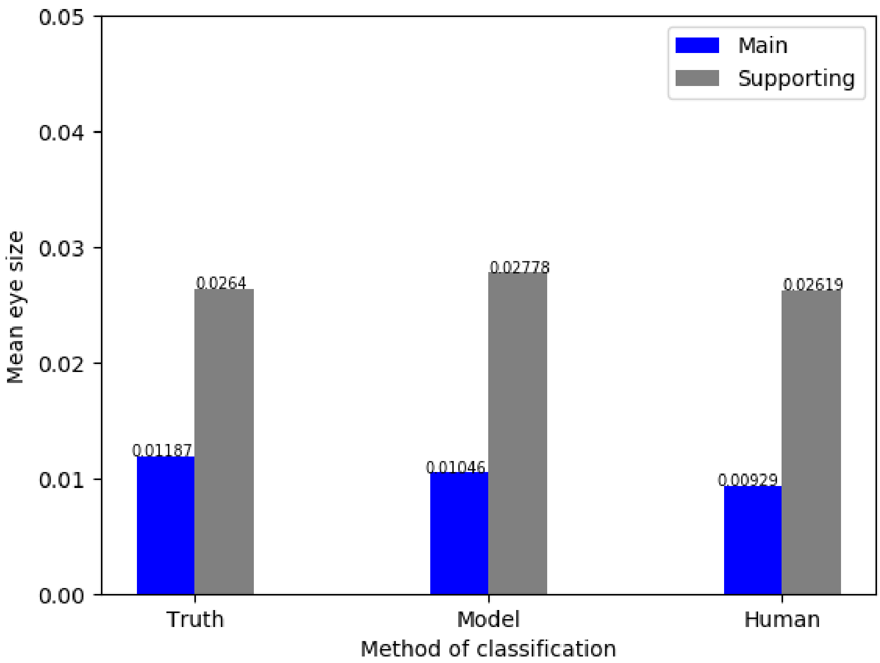


Figure 7.3: Graph showing how eye size is impacted by class.

7.5 Hair

As a visual feature, we'll divide hair by analysing its length and color separately. Both aspects play an important role in a character's overall design

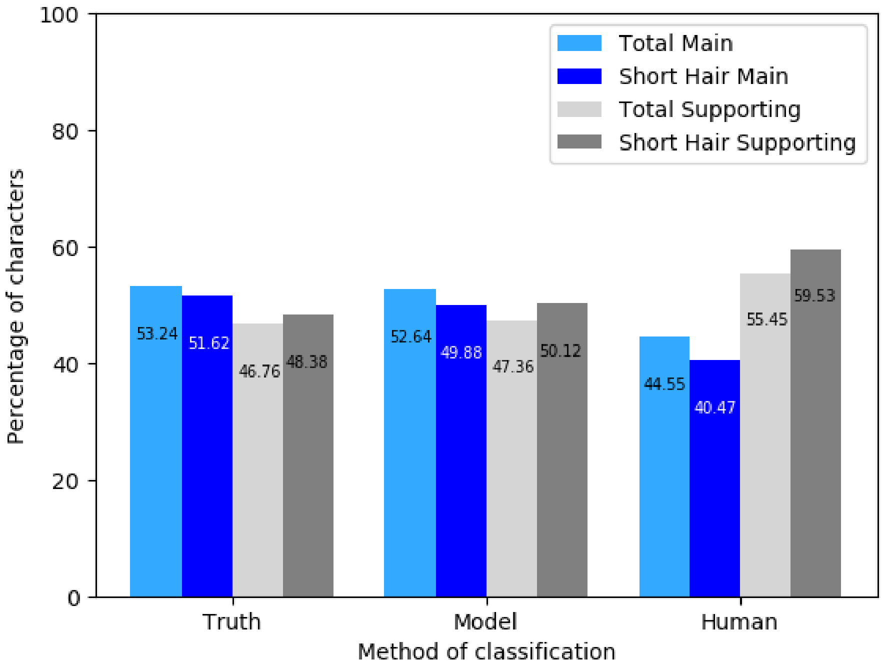
(28) as the feature is one of the main ways to differentiate between characters (28). Long hair can be associated with feminine beauty, and freedom (28) (29), thus supposedly lending itself well to female main characters (11), as these are characteristics usually associated with the role (10) (6). At the same time, color carries strong meaning in the medium (23), with the hair being one of the main elements used to convey it (29) (30).

7.5.1
Hair Length

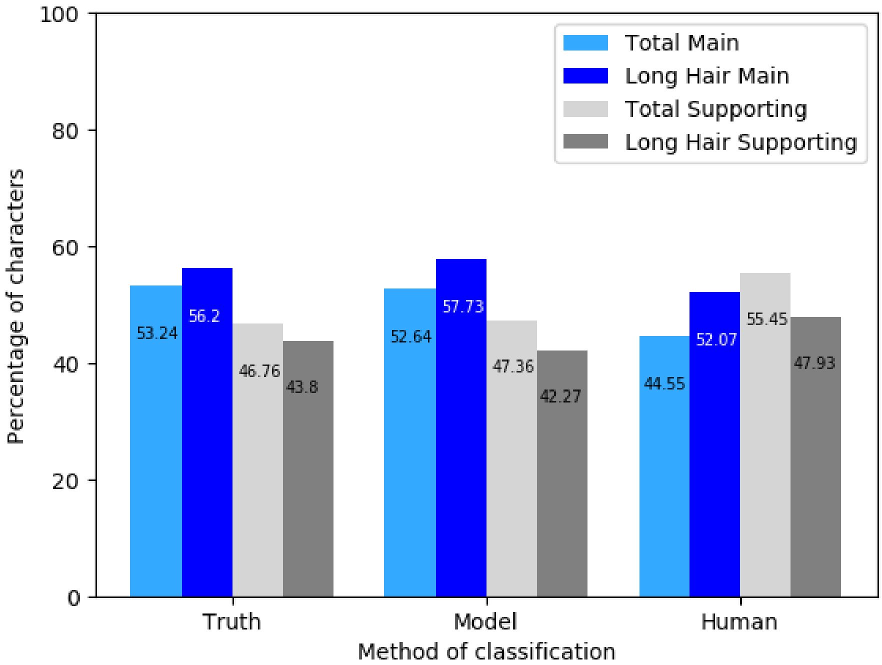
When considering all characters combined, a majority (64.78%) had short hair, but doing so represented a 3.04% lower chance of being a main character, with the model increasing the difference to 5.24% and the humans taking it even further to 9.15%. Those featuring long hair being 5.55%, 9.66%, and 16.87% more likely to be labeled as main characters, according to the ground truth, the model, and the humans, respectively. This indicates that, overall, characters with long hair are slightly more likely to be main characters, but it also shows that audiences tend to dramatically overestimate the importance of this factor, while the model shows its usual trend of exaggerating the existing patterns. However, this feature is heavily influenced by the character's apparent gender (11) (28) (29), so looking at each one in isolation may lead to more conclusive results.

Female characters were more likely to feature long hair in general, with 55.51% of them being part of that group. However, having long hair only represented a 3.97% higher likelihood of being a main character, when compared to the average female of characters, with the model showing a difference of only 1.3% and humans giving it the most importance at 9.91%. Reaffirming that, while there is a correlation between having long hair and being a main character, audiences have a tendency to overestimate its relevance as they attribute too much importance to the meaning behind hair design (29).

Male characters paint an even clearer picture of that tendency, with the presence of long hair representing a `14.18%` lower chance of being a main character, which reflects how, in contrast to its meaning in female characters, long hair in male characters can convey a violent and brutal nature (28), which are traits associated with antagonistic roles (10) (5) (6). However, the model and humans show a trend in the opposite direction, with the feature increasing the likelihood of one being a main character by `15.58%` and `8.10%`, respectively. This indicates that long hair, even in male characters, is still associated with the meaning commonly attributed to female characters, and the personality traits associated with it (28) (29), thus the feature acts as a strong indicator for audiences that a character is among the main characters, even if its presence is at best slightly relevant and, in the case of male characters, shows a trend in the opposite direction.



(a) Short Hair class distribution, when compared to all characters



(b) Long Hair class distribution, when compared to all characters

Figure 7.4: Graphs showing how hair length impacts class distribution.

7.5.2
Hair Color

The impact each hair color has on a character's likelihood of being part of the main cast can be seen in figure 5. The first notable fact is that grey hair represents the lowest likelihood of being a main character as the color is associated with old age and those characters are rarely put in leading roles.

Characters with purple, yellow, pink, blue, red, and orange hair were considered by all three classification methods to have a higher-than-average likelihood of being part of the main cast.

Blonde characters were considered more likely to be main characters regardless of gender, matching the expectations for the color as one that symbolizes uniqueness and being special (28) (30) due to its rarity in Japan. The color can also be associated with confidence, happiness and naivité (29), traits linked to the role of hero (10), especially early on their journey (5) (6), reinforcing its presence as an identifier of main characters. While the color is a relevant indicator of a character's importance, with blonde characters being 7.72% more likely to be main characters, audiences again overestimated its impact, considering those featuring it to be 28.72% more likely to be part of the main cast, reflecting the cultural expectations on the color detailed before. Meanwhile, the model remained close to the actual pattern with a higher likelihood of 8.93%. The color is also given special importance in the Shojo demographic as an indicator of being a protagonist (30), and while it's a strong indicator for female characters, both the actual data and the human survey results found it to have a stronger correlation with being a main character in the case of male characters.

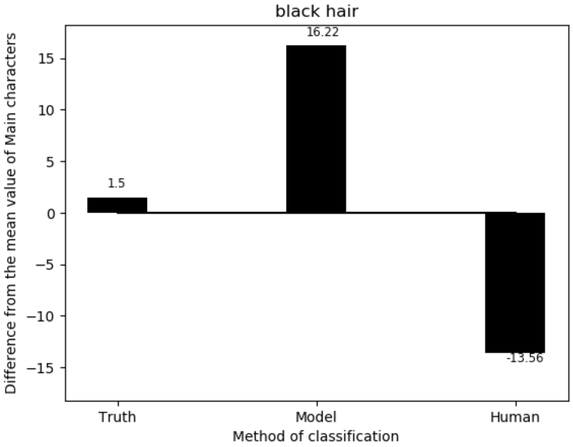
Characters with red and orange hair were found to have a higher than average likelihood of being main characters according to all classification methods and applying to both genders. Both colors carry similar meanings, usually associated with passion, leadership, violence and strong emotions more generally (28) (30), traits can be linked to the role of hero(6). And while the same colors can mean the opposite when it comes to some male characters, signifying that they're calm, humble, and disciplined (29), these traits can be associated with the lancer or ally (10) (6), making them also part of the main cast.

Characters with purple and blue hair served as good indicators of a character's importance for both genders, but while all classification methods agreed on the latter, the model showed the opposite pattern when it came to the former. Purple hair tends to mean power or privilege, as well as intelligence and mystery (28) (29) (30), traits that are usually associated with antagonistic or neutral roles (10) (5) (6), but they can also be present in archetypes potentially present in the main cast, such as the femme fatale (6). In addition, the color, especially in lighter tones, can also be linked to being magical or divine (29) (30), complementing the aspects of intelligence and mystery to form the role of boom giver (5, 7), or detached love interest (29), which can both be part of the main cast. Meanwhile, blue hair usually relates to being reliable, intelligent, calm, wise and confident (29) (30); the lack of negative associations with color means that it is not commonly used for antagonistic roles, while the fact that it's linked to being attractive (29) (30) keeps it from being used in old mentors who rarely are depicted as such (10) (6), thus explaining why the color is a strong indicator of a character being part of the main cast.

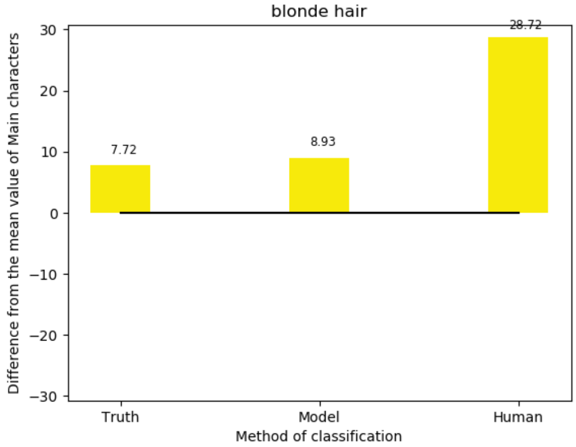
In female characters pink acted as the strongest indicator of being in the main cast. This held true for all classification methods. The color is mostly associated with youth, innocence, femininity, naivité and benevolence (28) (29) (30), traits that can be associated with some types of heroes (10) and made it ideal for the "moe-protagonist" archetype (30). However, in male characters the color can represent malicious or wicked characters (29), explaining why it beats only grey as an indicator of being a male protagonist.

For male characters, black also acted as a good indicator of being a main character, a pattern that the model followed but the humans didn't, possibly due to black being a very common color for hair (28) (30), losing only to brown in our data, and its common usage without narrative meaning (30). However, the pattern is real, with male characters with black hair being 20.01% more likely to be part of the main cast than the average of male characters. The color can be associated with being determined, independent, mysterious, powerful, practical, solitary (29) and despite being dark, it usually carries positive connotations (29), making it ideal for the lone wolf type of character (29), that can play the role of hero in many narratives (6). This type of character is more commonly associated with male characters, thus explaining the color's presence here, but not in female main characters.

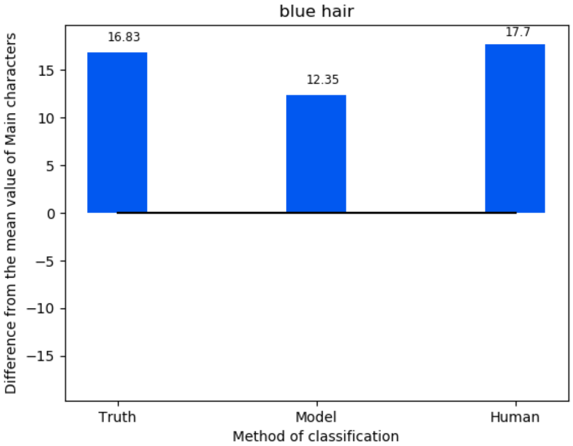
Other colors such as green and white represent a lower likelihood of being a main character regardless of gender, but humans judged the opposite in the case of both male and female characters with white hair. The unnatural color can relate to being lucid, skillful and serene, leading to female magical characters and male antiheroes (29), and these traits can be linked to protagonistic roles (6), possibly explaining, alongside its rarity, why audiences tend to attribute such importance to the color. However, having white hair can also mean a detachment from reality (30), which may turn antiheroes into villains (6), removing them from the main cast of protagonists. Differentiating between these two groups purely on their design may have proven difficult for humans, who ended up giving more weight to the positive connotations of the color.



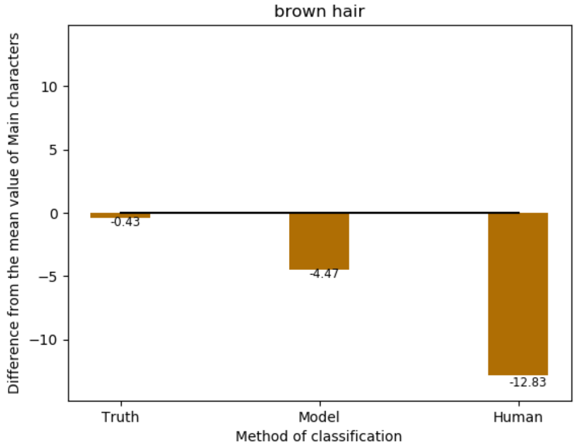
(a) Impact of black hair on the likelihood of being a main character



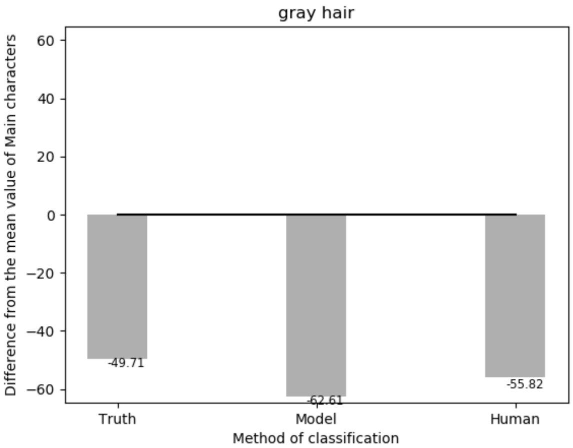
(b) Impact of blonde hair on the likelihood of being a main character



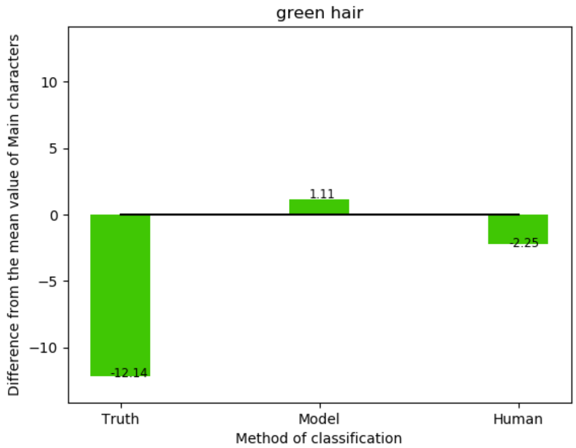
(c) Impact of blue hair on the likelihood of being a main character



(d) Impact of brown hair on the likelihood of being a main character

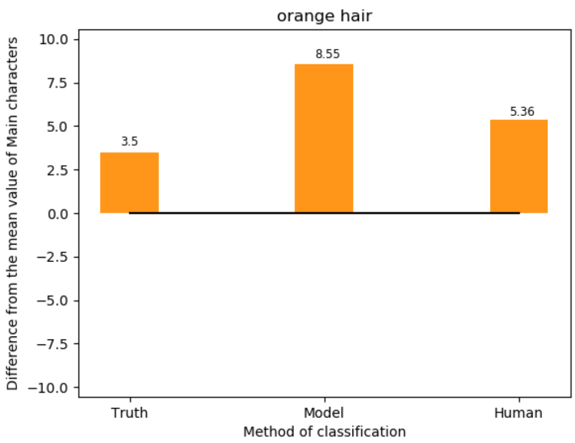


(e) Impact of gray hair on the likelihood of being a main character

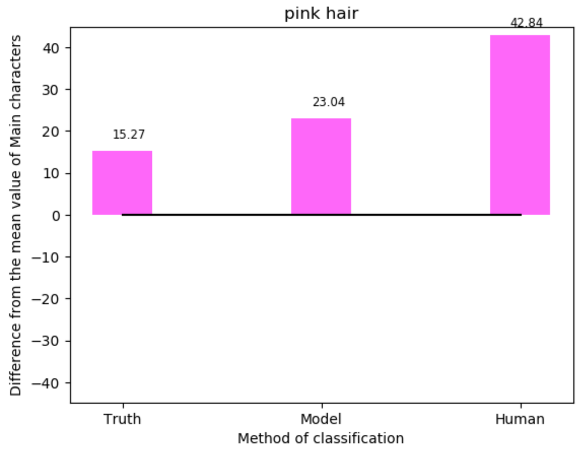


(f) Impact of green hair on the likelihood of being a main character

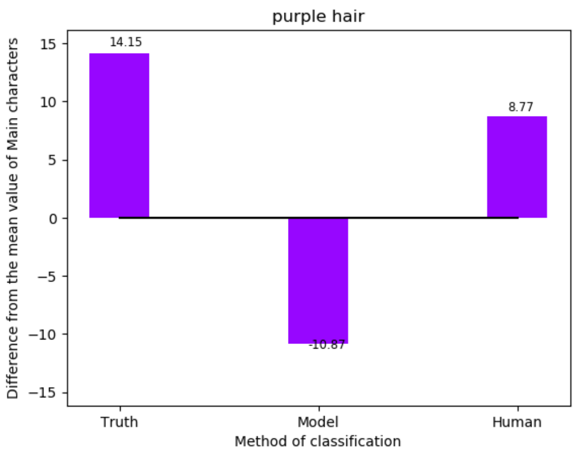
Figure 7.5: Graphs showing how hair color impacts the likelihood of being a main character.



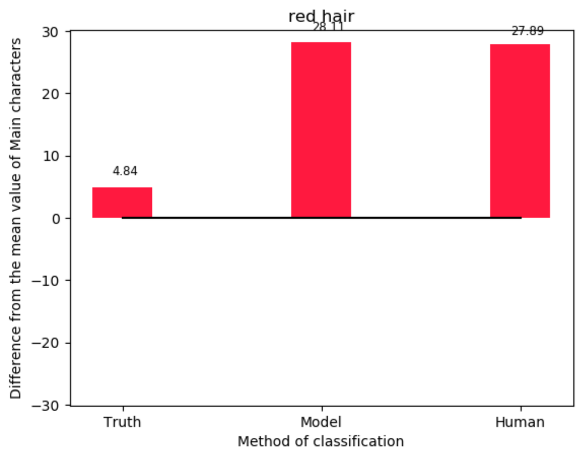
(a) Impact of orange hair on the likelihood of being a main character



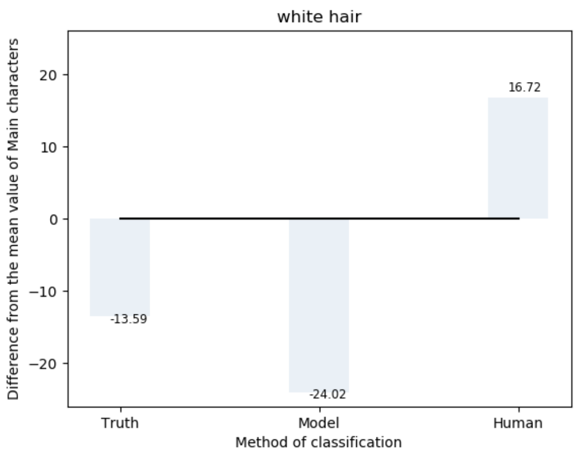
(b) Impact of pink hair on the likelihood of being a main character



(c) Impact of purple hair on the likelihood of being a main character



(d) Impact of red hair on the likelihood of being a main character



(e) Impact of white hair on the likelihood of being a main character

Figure 7.6: Graphs showing how hair color impacts the likelihood of being a main character.

8
Contributions

We aimed to help evaluate whether or not CNNs, especially deep residual architectures, are effective in tasks with slight inter-class variation and considerable intra-class variation, in particular those where classification can carry subjective elements, something especially true in tasks relating to person perception (4).

We studied a problem with little representation in the literature, possibly serving as start for future work on tasks of this nature.

We were able to show that, when it comes to the task we studied, CNNs can learn many of the same patterns shown by humans, even if some of those patterns don't reflect reality. Therefore, it means that this kind of model can be used to find the implicit patterns and biases humans use when labelling objects in tasks of this nature, and also that the technology might be a suitable representative of human classifiers in areas such as image generation for this kind of task.

We also contributed by creating a dataset for our specific task, and providing a performance baseline on it, making any further work on the task easier and quicker, as well as give it a better sense of direction since there will be a point of comparison.

8.1
Future Work

To continue the work done here, it might be worth testing other uses of the attention mechanism, especially during feature extraction, as it may improve the model's ability to evaluate different features.

When it comes to the specific task we studied, it has been shown that covariance pooling (31) can be used to improve performance in extracting features from anime faces (14) due to its ability to better evaluate distortions in facial features. While these distortions aren't as directly linked to narrative importance as they were with emotion detection, the technique could still prove useful for our task and is worth considering.

9
Bibliography

1 WEI, X.-S. et al. **Fine-Grained Image Analysis with Deep Learning: A Survey**. arXiv, 2021. Disponível em: <https://arxiv.org/abs/2111.06119>. Cited in page 13.

2 DENG, J. et al. Imagenet: A large-scale hierarchical image database. **IEEE conference on computer vision and pattern recognition**, may 2009. Cited 2 times in pages 13 and 24.

3 KRIZHEVSKY ILYA SUTSKEVER, G. E. H. A. Imagenet classification with deep convolutional neural networks. **NIPS**, 2012. Disponível em: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. Cited 3 times in pages 13, 14, and 15.

4 CANTOR, W. M. N. **Advances in experimental social psychology**. [S.l.]: Elsevier Inc., 1979. Prototypes in person perception. Cited 3 times in pages 13, 18, and 45.

5 CAMPBELL, J. **The Hero with a Thousand Faces**. 3. ed. [S.l.]: New World Library, 2012. Cited 7 times in pages 13, 16, 19, 32, 38, 40, and 41.

6 VOGLER, C. **The Writer's Journey**. 3. ed. [S.l.]: Michael Wiese Productions, 1998. Cited 10 times in pages 13, 16, 17, 32, 35, 37, 38, 40, 41, and 42.

7 PROPP, V. **Morphology of the Folktale**. 1. ed. [S.l.]: University of Texas Press, 1968. Cited 3 times in pages 13, 17, and 41.

8 HOFFNER, J. C. C. Developmental differences in responses to a television character's appearance and behavior. **Developmental Psychology**, v. 21, n. 6, 1985. Cited 2 times in pages 14 and 17.

9 HOFFNER, J. C. C. Responding to the screen: Reception and reaction processes. In: _____. [S.l.: s.n.], 1991. cap. Perceiving and responding to mass media characters, p. 63–101. Cited 2 times in pages 14 and 17.

10 ROGERS MARIA AUFHEIMER, M. W. L. E. N. K. Towards the visual design of non-player characters for narrative roles. **Graphics Interface Conference**, may 2018. Disponível em: <https://graphicsinterface.org/wp-content/uploads/gi2018-21.pdf>. Cited 12 times in pages 14, 16, 17, 19, 21, 32, 33, 35, 37, 38, 40, and 41.

11 LIU, P. W. M. Study on image design in animation. **Asian Social Science**, v. 6, n. 4, apr 2010. Disponível em: <https://pdfs.semanticscholar.org/6202/af33ab7918cafa3622010e62d245f12d0713.pdf>. Cited 9 times in pages 14, 17, 18, 20, 22, 30, 33, 36, and 37.

12 HE XIANGYU ZHANG, S. R. J. S. K. Deep residual learning for image recognition. **IEEE conference on computer vision and pattern recognition**, dec 2015. Disponível em: <https://arxiv.org/pdf/1512.03385.pdf>. Cited 2 times in pages 15 and 24.

13 LEDIG LUCAS THEIS, F. H. J. C. A. C. A. A. A. A. A. T. J. T. Z. W. W. S. C. Photo-realistic single image super-resolution using a generative adversarial network. **IEEE conference on computer vision and pattern recognition**, may 2017. Disponível em: <https://arxiv.org/pdf/1609.04802.pdf>. Cited in page 15.

14 JESUS JESSICA CARDOSO, A. B. G. S. C. R. M. I. A cnn-based tool to index emotion on anime character stickers. **2019 IEEE International Symposium on Multimedia (ISM)**, december 2019. Disponível em: <https://www.researchgate.net/profile/Antonio-Busson/publication/338651794_A_CNN-Based_Tool_to_Index_Emotion_on_Anime_Character_Stickers/links/5fc40b8b458515b79788d5af/A-CNN-Based-Tool-to-Index-Emotion-on-Anime-Character-Stickers.pdf>. Cited 2 times in pages 15 and 45.

15 JIN JIAKAI ZHANG, M. L. Y. T. H. Z. Z. F. Y. Towards the automatic anime characters creation with generative adversarial networks. **Comiket 92 (Summer 2017), Tokyo Big Sight**, aug 2017. Disponível em: <https://arxiv.org/pdf/1708.05509.pdf>. Cited in page 15.

16 GOODFELLOW JEAN POUGET-ABADIE, M. M. B. X. D. W.-F. S. O. A. C. I.; BENGIO, Y. Generative adversarial nets. **Advances in neural information processing systems**, jun 2014. Disponível em: <https://arxiv.org/pdf/1406.2661.pdf>. Cited in page 15.

17 PRINCE, G. **A dictionary of narratology**. 1. ed. [S.l.]: University of Nebraska Press, 1989. Cited in page 16.

18 RADFORD JONG WOOK KIM, C. H. A. R. G. G.-S. A. G. S. A. A. P. M. J. C. G. K. I. S. A. Learning transferable visual models from natural language supervision. feb 2021. Disponível em: <https://arxiv.org/pdf/2103.00020.pdf>. Cited 3 times in pages 18, 19, and 20.

19 LIU JUN-HONG CHEN, K.-M. C. K. A study of facial features of american and japanese cartoon characters. **Symmetry**, may 2019. Disponível em: <https://www.mdpi.com/2073-8994/11/5/664/htm>. Cited in page 19.

20 VALENTIN. **Why do Anime Characters have Big Eyes?** 2013. Disponível em: <https://craftknights.com/why-do-anime-characters-have-big-eyes-the-history-of-the-anime-style/#:~:text=Why%20are%20Anime%20Eyes%20so,of%20the%20Character%20to%20much>. Cited 2 times in pages 20 and 35.

21 LIU KANG-MING CHANG, Y.-J. L. J.-H. C. K. Animated character style investigation with decision tree classification. **Symmetry**, jul 2020. Disponível em: <https://www.mdpi.com/2073-8994/12/8/1261/pdf>. Cited in page 20.

22 SAITO, Y. M. M. Illustration2vec: A semantic vector representation of illustrations. **SIGGRAPH Asia 2015 Technical Briefs**, 2015. Disponível em: <https://www.gwern.net/docs/anime/2015-saito.pdf>. Cited in page 20.

23 CHENG, W. Analysis on nationality of japanese anime color. **IEEE 10th International Conference on Computer-Aided Industrial Design Conceptual Design**, nov 2009. Cited 2 times in pages 20 and 37.

24 KABIR MOLOUD ABDAR, S. M. J. J. A. K. A. F. A. S. N. D. S. H. M. D. Spinalnet: Deep neural network with gradual input. jul 2020. Disponível em: <https://arxiv.org/pdf/2007.03347.pdf>. Cited in page 24.

25 ZHENG, M. et al. Attention-based cnns for image classification: A survey. **Journal of Physics: Conference Series**, IOP Publishing, v. 2171, n. 1, p. 012068, jan 2022. Disponível em: <https://dx.doi.org/10.1088/1742-6596/2171/1/012068>. Cited in page 25.

26 CATEGORY:ANIME and manga by demographic. 2022. Disponível em: <https://en.wikipedia.org/wiki/Category:Anime_and_manga_by_demographic>. Cited 2 times in pages 29 and 32.

27 HICKSON, T. **On Writing and Worldbuilding**. [S.l.: s.n.], 2019. v. 1. Cited in page 32.

28 FOX, A. **A Viewer's Guide to Anime Hair's Meaning**. 2013. Disponível em: <https://www.forevergeek.com/anime-hairs-meaning/>. Cited 5 times in pages 37, 38, 40, 41, and 42.

29 KEVIN. **Anime Hair - Colors and Hairstyles and their Meanings**. 2017. Disponível em: <https://skdesu.com/en/hair-in-anime-colors-and-hairstyles-and-their-meanings/>. Cited 5 times in pages 37, 38, 40, 41, and 42.

30 SCHNEIDER, M. **Anime Hair Colors: Do They Carry Any Significant Meaning In Japanese Culture?** 2015. Disponível em: <https://thoughtcatalog.com/martin-schneider/2015/02/anime-hair-colors-do-they-carry-any-significant-meaning-in-japanese-culture/>. Cited 4 times in pages 37, 40, 41, and 42.

31 LIU, M. et al. Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild. **Proceedings of the 16th International Conference on Multimodal Interaction**, 2014. Cited in page 45.