Visual Foundation Model-Based Classification of Characters in Narrative Media

Yan Martins B. Gurevitz Cunha
Telemidia Lab. – PUC-Rio
yangurevitz@telemidia.puc-rio.br

Daniel de Sousa Moraes
Telemidia Lab. – PUC-Rio
danielmoraes@telemidia.puc-rio.br

Antonio José Grandson Busson
Telemidia Lab. – PUC-Rio
busson@telemidia.puc-rio.br

Julio Cesar Duarte
Military Institute of Engineering
duarte@ime.eb.br

Sérgio Colcher
Telemidia Lab. – PUC-Rio
colcher@inf.puc-rio.br

ABSTRACT

The field of narratology has long worked on different ways to classify characters according to their narrative role and importance; meanwhile, studies on character design have observed the existence of unclear patterns that correlate specific visual features and those narrative classifications. Therefore, it becomes worth considering if modern image classification methods can be used to not only achieve satisfactory performance on this task, but also help us learn about the patterns that shape it. The task of classifying characters carries a vast array of practical uses, such as creating recommender systems based on specific visual-narrative correlations, automatic annotation for media preservation, adaptive interactive storytelling based on how a character is perceived, and the creation of education tools on how those patterns change depending on culture. In this work, we construct a dataset customized to assessing both performance and visual–narrative correlations within a specific medium, conduct an anonymous survey to evaluate human performance and tendencies, and compare the results with those of a traditional CNN classifier and a Foundation Model–augmented approach. Both models outperform the survey participants, with the latter achieving 76.93% accuracy, 18.11% higher than human performance, while also providing insights into the visual cues most closely tied to the given narrative roles.

KEYWORDS

Image Classification, Deep Learning, Anime Characters, Narratology, Foundation Model

1 INTRODUCTION

Classifying characters based on their visual appearance is a task with far-reaching implications in areas such as media analysis, entertainment, human–computer interaction, and digital content management. In streaming platforms, for example, automated character recognition can enable advanced search and recommendation systems that identify recurring characters across series or franchises, even when portrayed in different artistic styles, as well as facilitating archiving and media restoration in animation by speeding up indexing and retrieval in large-scale visual archives [17]. In the gaming industry, adaptive storytelling systems could adjust dialogue, difficulty, or narrative progression based on how non-player characters perceive the player character [32]. Beyond entertainment, such models could also serve as educational tools that teach narrative structure or as a basis for sociocultural studies analyzing how character visual design reflects and influences audience perceptions across cultures [24].

In narrative media, a character's visual design is deliberately crafted to convey personality traits, social roles, and narrative importance, subtly guiding audience expectations [5, 6]. However, these cues are rarely explicit and can vary significantly across cultural contexts, artistic conventions, and individual interpretations, making classification inherently challenging [25]. Unlike traditional image classification tasks—such as those exemplified by ImageNet [7]—the differences between categories in character classification are often subtle, with minor inter-class variations and substantial intra-class variability. This requires both humans and machine learning models to interpret ambiguous and sometimes misleading patterns, a challenge amplified when distinctions concern abstract narrative roles, such as differentiating a protagonist from a supporting character. Addressing this challenge not only advances computer vision applied to narrative media but also deepens our understanding of how visual cues influence perception and decision-making in broader contexts.

Tasks where visual cues are subtle or non-discriminative remain underexplored in the literature, creating a gap in our understanding of model performance under such constraints. While traditional architectures, such as Convolutional Neural Networks (CNNs)[1], have been widely used, the emergence of visual foundation models, including CLIP[29] and DINOv2 [26], offers new opportunities in this domain. These models are trained on massive and diverse datasets, capturing generalizable visual representations that can be adapted to niche tasks with limited labeled data. Their ability to use both global semantic information and fine-grained visual details makes them promising candidates for classification scenarios where the visual distinctions are subtle and context-dependent, as is the case in determining narrative importance.

To address this gap, this paper focuses on a specific and challenging instance of subtle-pattern classification: predicting the narrative importance of a character solely from its visual appearance. In narrative theory, a character's role is often tied to their influence on the storyline [2, 28, 34]. Prior research shows that visual design strongly influences audience expectations [5, 6], with specific design traits correlating with particular narrative roles [15]. As Rogers

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2025). Rio de Janeiro, Brazil. Porto Alegre: Brazilian Computer Society, 2025.

ISSN 2966-2753

et al. [15] note, while individual interpretations may vary, there tends to be consensus regarding a character's perceived narrative status based on visual design, suggesting a learnable mapping between appearance and role.

The objectives of this work are therefore twofold: (1) Evaluate and compare the performance of CNN-based classifiers and visual foundation models in predicting whether a character is a main or supporting figure based solely on imagery; and (2) Identify visual patterns in character design that correlate with narrative importance, offering insights into both the specific problem and the broader challenge of subtle-pattern image classification.

To this end, we propose and release a dataset of anime characters, chosen for their visual consistency, stylized traits [24], and relative simplicity compared to photorealistic imagery. We establish a human performance baseline via an anonymous survey and compare it with CNN and foundation model baselines, analyzing the strengths and weaknesses of each.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature in image classification, narrative theory, and foundation model applications. Section 3 presents the dataset creation process and the models used. Section 4 describes the experimental setup and compares results across methods. Finally, Section 5 discusses the implications of our findings, their relevance to the broader problem of subtle-pattern classification, and potential avenues for future work.

2 RELATED WORK

This section reviews previous work related to image classification involving ambiguous patterns, with a particular focus on the specific case of classifying narrative importance.

There were a few other works correlating character design and narrative role in narratology, and CNNs [1] have seen extensive study in the past decade. There have also been studies on character generation [13]. 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.

2.1 Image Classification

Gomes et al. [9] have demonstrated how attention features extracted from self-supervised foundation models can be combined with traditional residual networks to achieve greater performance in tasks involving human image classification. This raises the question of whether that approach could also be applied to a task where the different classes are more subtle and ambiguous. No prior research was found that combines character design and narrative role within deep learning-based classification and data analysis, so we relied on foundational works from each respective field.

We have also seen image classification techniques being applied to the context of narrative media, such as Kim et al. [17], who showed considerable progress in the task of character detection for animated works by employing a multi-layered network that manages to distinguish between different visual styles and human and non-human characters. Those gains were valuable to the field of character detection. Still, the same approach wouldn't suit our goals since the visual patterns that determine whether a character is main or supporting are not as clear as those that identify that character as its individual.

Another example is the work by Laubrock and Dubray [19], who explored the use of CNNs to classify illustrator styles in graphic novels, aiming to identify which visual features contribute most to distinguishing between artists. The authors train a CNN on a dataset of graphic novel panels from different illustrators. Their findings reveal that mid-level features, such as texture and stroke patterns, are particularly discriminative for illustrator style. In contrast, low-level features (edges, colors) and high-level semantic content play a lesser role. The study highlights the potential of CNNs for art historical and stylistic analysis while providing insights into how neural networks perceive artistic style.

Regarding works dealing with human faces in multimedia environments, Mendes et al. [23] proposed an approach for video-based face recognition that uses cluster matching to improve accuracy in unconstrained environments (e.g., varying poses, lighting, and occlusions). They grouped face images from videos into clusters based on visual similarity and then matched entire clusters between reference and probe videos. Experiments demonstrate that the cluster-matching approach outperforms single-image and temporal pooling methods. The key advantage lies in its ability to discard outliers and focus on stable, high-confidence face representations within clusters. The paper also discusses computational efficiency, making it practical for real-world applications like surveillance and video indexing. The results suggest that cluster-based matching is a promising direction for improving video face recognition systems.

And Jesus et al. [12] used covariance pooling and a ResNet architecture to label the faces of anime characters according to their emotions. Their work also involved creating an elaborate dataset exclusive to the task from images available online and processing it to ensure focus on the facial features. The considerable performance they were able to achieve on the task shows how traditional classifiers can perform in tasks relating to medium and how its general visual simplicity can be an asset when attempting to learn its patterns.

2.2 Narratology

First, it is crucial to establish the concept of narrative role we will be working with, as established by Prince [27], who defines it as a set of functions that can be performed by a narrative entity (usually, but not limited to, a character). 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. That definition is then expanded on by Vogler [34], who established a correlation between narrative role and narrative importance, affirming, for instance, that a hero is more likely to be a main character than a trickster. That relationship between narrative role and expected prominence will be significant when analyzing the patterns we encountered.

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 [2], their refinement by Vogler [34], and Propp's dramatis personae [28].

Despite their differences, these frameworks share specific core roles, such as the hero and the villain (or shadow).

2.3 Character Design

In the field of character design, we have seen attempts to correlate visual and narrative elements in different types of media. For instance, Rogers et al. [15] investigated video game Non-Player Characters, which helped find correlations between certain visual traits and their expected narrative role. Hoffner and Cantor [5, 6] examined 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 [24] analyzed visual design in animation, focusing specifically on how Japanese animation establishes distinctive traits that signal main characters. Having these correlations between visual features and narrative role, alongside the previously mentioned correlations between narrative role and importance, allows us to establish a base of understanding from which we can interpret the results of our experiments. These works, unlike our proposal, aimed to evaluate the audience's expectations and did not measure how well those match reality when it comes to a large set of pre-existing characters.

3 DATASET

To support the evaluation of different classification approaches in the character classification task, we developed a dedicated dataset composed of anime character images paired with reliable narrative prominence labels. Anime character imagery is widely available online, with several repositories containing images for more than 100,000 characters. For this work, we focused on MyAnimeList (MAL)[1] as our primary source, since the platform not only provides a vast and diverse collection of characters but also categorizes each one as Main or Supporting, based on their role in the narrative. This feature allowed us to obtain trustworthy ground-truth labels for supervised learning. Using this approach, we compiled an initial collection of over 31,000 characters, each represented by a single image.

To ensure the dataset aligned with our analytical objectives, we applied a filtering step using CLIP [29]. This automated classification was employed to exclude characters whose appearance was predominantly animalistic, fantastical, or mechanical. The rationale for this filtering is that such traits may influence the perception of the characters in ways that differ from human-focused cues, which are central to our study. By retaining only those labeled as human, we aimed to reduce variability that could obscure the evaluation of narrative importance models.

This filtering step was particularly relevant for employing theories from person perception [25], which emphasize the role of human physical and facial features in social and narrative interpretation. By focusing on human characters, the dataset allows for more direct investigation into how visual cues relate to narrative prominence. As a result, the constructed dataset provides a robust and well-curated resource for future studies on visual character analysis, narrative importance prediction, and other tasks involving human-centered image understanding in anime media. Figure Figure 1 shows a few examples of characters belonging to different classes, even though they share many visual features.

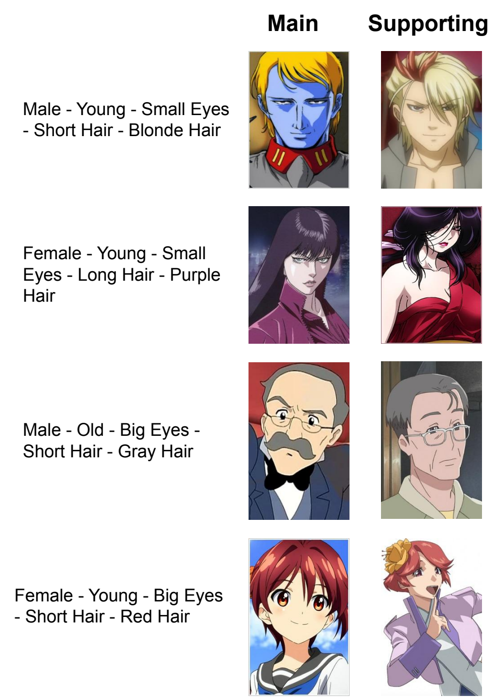

Figure 1: Examples of characters belonging to different classes while still sharing visual features. Source: https://myanimelist.net/

3.1 Apparent Gender

Although a character's true gender identity within the narrative may differ from their appearance, for this task, we limited ourselves to their apparent gender as classified by CLIP [29], since we are working with only their character design without external narrative context. This left us with 45.20% of characters classified as `male` and 54.80% as `female`. This 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.

3.2 Apparent Age

Similarly to gender, a character's actual age may differ from their appearance. We still focused on their apparent age, as this was the only information available to both the model and the survey participants. Here, the division was considerably 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

[1] https://myanimelist.net/

given that audiences tend to associate old age with few narrative roles beyond that of the mentor [15], which can even be seen in Campbell's original name for the archetype as the Wise Old Man or Woman [2].

3.3 Eyes

Exaggerated eyes are an essential element in both American and Japanese animation, with the latter showing greater intensity and more frequent use of this concept [18]. 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 [33]. Therefore, it is not surprising that their size would be an essential factor in defining the appearance of the main characters [24]. The fact that the use of this feature in anime originated in Shojo (teenage girls) manga [20], alongside the importance Minghua Liu and Ping Wang [24] place on it specifically for female main characters, acts as a strong indication that a character's gender may heavily influence this visual trait. To quantify eye size, we used Illustration2Vec's tag extractor [22], interpreting the model's likelihood of a character having closed eyes as an inverse indicator, with higher values corresponding to smaller eyes. Main characters are, on average, half as likely to have closed eyes compared to supporting characters.

3.4 Hair

As Minghua Liu and Ping Wang [24] note, hair plays a vital role in anime character design, especially in the case of female main characters. Considering also how color is a key factor in conveying narrative meaning in the medium [3], we measured how aspects 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 [29] 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. The influence of this feature on narrative classification for male and female characters is discussed when we analyze the results of our experiments in Section 4. The distribution of hair colors was as follows: black (19.04%); blonde (11.16%); blue (7.75%); brown (28.27%); gray (6.24%); green (3.80%); orange (3.28%); pink (4.16%); purple (6.68%); red (5.10%); and white (4.52%). The fact that over 60% of characters feature usual hair colors such as black, brown, blonde, and gray shows an interesting contrast to anime's reputation for featuring characters with unusual and intense hair colors.

4 EXPERIMENTS AND ANALYSIS

To compare the performance of different classification methods on the task and the patterns they exhibit in their correlation between visual features and narrative importance, we established three primary classification methods that covered a variety of different approaches. Firstly, we conducted a human survey to measure baseline performance on the task and observe which patterns our participants would show regarding the correlation between visual features and their classification. Subsequently, we trained both a traditional CNN classifier and one that makes use of attention maps from foundation models to compare their performance and data pattern recognition against the survey results.

The remainder of this section is structured as follows. Subsection 4.1 goes over the survey conducted for this research, its process, and results. Subsection 4.2 details the classifiers tested, their performance, and how their results compare to the human baseline. And Table 1 summarizes the performance metrics for each classification method tested.

4.1 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 human performance on the task of classifying a character between main and supporting roles based solely on their images. This measurement serves as a baseline for comparing the performance of our models.

Our initial expectations were mixed. On one hand, previous studies have suggested that audiences are better at identifying a character's role than at determining which visual features indicate that role [15]. On the other hand, the correlations between visual and narrative features in our data were unclear, making us unsure of which patterns survey takers would rely on. Therefore, in addition to quantitative performance, we were also interested in investigating the biases influencing participants' judgments.

The survey was conducted online with people familiar with the medium of anime. Each participant was presented with one character at a time and had to classify each one according to whether they thought it was a main or supporting character. They were also given the option to indicate if they already knew the character, ensuring that valid classifications were based solely on the image presented. The characters shown to survey takers were pulled from the test subset of the dataset and were shown in a different random order to each survey taker. Each participant could classify as many characters as they wanted, with the results being made completely anonymous before saving.

In total, 21 participants provided over 6,000 individual classifications, possibly offering 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.

Test takers achieved an accuracy of 58.82% compared to the ground truth. This suggests that while people are somewhat able to determine how important a character is at first glance, their performance may be hindered by biases, either due to simply unfounded preconceptions or deliberate subversion 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, our participants classified only 44.55% of them as main and 55.45% as supporting. This indicates that humans may have more demanding expectations of what main characters should look like, a notion reinforced by the relatively strict definition of the role provided by Minghua Liu and Ping Wang [24], as well as the specific expectations of each visual element, which are reflected in the comparison of these results with our models.

4.2 Classifiers

4.2.1 Traditional Classifier. 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 Section 3, keeping the same subset used for the human survey as the testing subset.

The models selected for this study primarily included various ResNet architectures [14].We alternated between training models from scratch and fine-tuning pre-trained models, while at the same time testing different hyperparameters such as optimizers, learning rates, and data augmentation techniques. We found that larger models quickly overfit, requiring low learning rates to avoid this problem. This issue can be caused by the indirect relationship between visual features and labels, a challenge that may also appear in other tasks of this nature.

Our initial best results were achieved with a ResNet50 model pre-trained on the ImageNet dataset [7] and fine-tuned using the RMSprop optimizer, reaching 64.56% accuracy on the test set, approximately 6% higher than the surveyed humans. We use this model as a basis for comparing patterns in visual feature recognition.

In contrast to how humans only classified 44.55% of the characters as main, the model closely matched the ground truth proportion of 53.24%, classifying 52.64% of characters as main and 47.36% as supporting. This can also be seen when considering the higher precision and recall, 66.90% and 61.95%, respectively. This suggests that the model may be considerably less influenced by the biases that affect human performance; however, a comparison of their results across each visual feature can provide a clearer understanding of this effect.

In most cases, the survey takers and the ResNet model agreed, 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 participants in 46.77%. The fact that the percentage of agreement between the two groups was higher than the model's accuracy reveals an interesting fact: the model was slightly better at mirroring human pattern perception than it was at learning the patterns actually present in the data.

Examining these numbers by class reveals clearer patterns. For supporting characters, the overall agreement rate was slightly higher at 68.74%, with both the model and participants correctly classifying 73.15% of cases. However, the most important aspect is when the two groups gave different answers, with the model being correct in only 39.85% of those cases, while participants were correct in 60.15%. This reflects the fact that, as previously mentioned, survey takers were more likely to classify characters as supporting when compared to both the model and the ground truth, making them less likely to mistakenly classify a supporting character as main. Conversely, for main characters, agreement occurred in only 64.84% of cases, and out of the ones they disagreed, the model was right in 63.68%, meaning humans were considerably more likely to misclassify a main character as supporting. This highlights the importance of identifying which visual factors played the most significant role in determining these classification differences. Table 2 shows the impact of different visual features on a character's likelihood of being classified as a main character.

4.3 Foundation Model Enhanced Classifier

We also tested how a more modern classifier performed on the task. To that end, we employed the approach proposed by Gomes et al. [9], which consists of using an attention map generated by a foundation model as an additional channel when training a CNN classifier. This technique has shown significant performance uplift and relatively low bias in a task related to human features, making it a good candidate for this task.

Following previous results obtained with this approach, the foundation model and classifier used were, respectively, DINOv2 [26] and XceptionNet [4]. This method achieved a performance gain greater than 12% over the ResNet model, with 76.93% accuracy. Due to its greater performance, this classifier will be our main focus when analyzing the correlation between visual features and the classification given.

4.3.1 Apparent Gender. When considering the distribution of classes according to apparent gender, 49.76% of male characters 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 [31], with different series targeting people of different genders and age groups.

The FM-enhanced classifier exacerbated the trend, with male characters being 9.26% less likely to be classified as main and female ones 8.18% more likely to receive the same label. Humans exhibited an even stronger version of the pattern, with males 11.23% less likely and females 9.36% more likely to be labeled as main characters.

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 participants giving the most importance to this feature. This factor must be carefully considered when analyzing the influence of other features, especially given that expectations of a character's appearance are heavily influenced by gender [24].

4.3.2 Apparent Age. When analyzing how apparent age impacts the classification of a character, it is important to remember that a large majority (87.43%) of characters were considered young, so the class distribution within this age group is most likely to be significantly closer to the overall distribution compared to that of older 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 it makes sense when we consider that old age is usually associated with wisdom and the role of mentor [15] and that these types of characters tend to take on supportive roles [2, 34], helping the protagonist and their close allies. Additionally, 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 [10]. 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 [31], with the first two being mostly focused on teen boys and girls, respectively, and the latter two on young men and women, respectively. Also, as the main characters are the ones the audience is supposed to more closely identify and relate with [34], it is natural that they are more likely to reflect their target demographic.

Both the model and the surveyed humans showed more intense versions of this trend, with the latter considering a character 67.54% less likely to be a main character if they were old, and the former again showing a more extreme result as it considered an old character to be 69.18% 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 the significant role a character's age plays in determining narrative importance, providing our first clear insight into the typical appearance of anime main characters, which 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 [15, 24] for both male and female main characters. Furthermore, these results also reinforce the general idea that these unclear correlations between visual features and classification tend to be exacerbated when trying to learn the task itself.

4.3.3 Eye Size. The expectations for the feature of eye size were that main characters would be more likely to have big eyes, since these are used to communicate a character's emotions [33] and the protagonists are the ones whose emotions and mental state tend to receive the most attention [34]. In addition, big eyes can be correlated with being good or innocent [33], traits associated with the role of hero [15].

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 is not as much of a distinguishing factor when it comes to female characters. However, the fact that both humans and the model did not 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. 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. While this means that stating that female protagonists tend to have big eyes [24] is not wrong, this feature alone is also not very helpful in finding a pattern for the role, since female characters generally 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 Minghua Liu and Ping Wang [24] suggest, 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 considerably more likely to be a main character, but having small eyes does not carry the same effect in the opposite direction.

4.3.4 Hair Length. As a visual feature, we treat hair by analyzing its length and color separately. Both aspects play an important role in a character's overall design [8] as the feature is one of the main ways to differentiate between characters [8]. Long hair can be associated with feminine beauty, and freedom [8, 16], thus supposedly lending itself well to female main characters [24], as these are characteristics usually associated with this role [15, 34]. At the same time, color carries strong meaning in the medium [3], with the hair being one of the main elements used to convey it [16, 30].

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 decreasing the difference to 2.13%, while the participants increased it to 9.15%. Those featuring long hair are 5.55%, 4.48%, 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 [8, 16, 24], 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 of female 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 tend to overestimate its relevance as they attribute too much importance to the meaning behind hair design [16].

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 [8], which are traits associated with antagonistic roles [2, 15, 34]. 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 [8, 16], 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.

4.3.5 Hair Color. The first notable fact regarding hair color is that gray 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 yellow, pink, blue, red, and orange hair were considered by all 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 [8, 30] due to its rarity in Japan. The color can also be associated with confidence, happiness, and naiveté [16], traits linked to the role of hero [15], especially early in their journey [2, 34], 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 9.01%.

Characters with red and orange hair were found to have a higher-than-average likelihood of being main characters according to all classification methods, a trend consistent for both genders. Both colors carry similar meanings, usually associated with passion, leadership, violence, and strong emotions more generally [8, 30], traits that can be linked to the role of hero [34]. And while the same colors can mean the opposite when it comes to some male characters, signifying that they're calm, humble, and disciplined [16], these traits can be associated with the lancer or ally [15, 34], 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 ResNet showed the opposite pattern when it came to the former, with the FM-enhanced model showing a very slight positive impact for the color. Purple hair tends to mean power or privilege, as well as intelligence and mystery [8, 16, 30], traits that are usually associated with antagonistic or neutral roles [2, 15, 34], but can also be present in archetypes potentially present in the main cast, such as the femme fatale [34]. In addition, color, especially in lighter tones, can also be linked to being magical or divine [16, 30], complementing the aspects of intelligence and mystery to form the role of boom giver [2, 28], or detached love interest [16], which can both be part of the main cast. Meanwhile, blue hair usually relates to being reliable, intelligent, calm, wise and confident [16, 30]; the lack of negative associations with the color means that it is not commonly used for antagonistic roles, while the fact that it is linked to being attractive [16, 30] keeps it from being used in old mentors who rarely are depicted as such [15, 34], thus explaining why the color is a strong indicator of a character being part of the main cast.

In female characters, pink hair acted as the strongest indicator of being in the main cast, consistent across all classification methods. The color is mostly associated with youth, innocence, femininity, naiveté, and benevolence [8, 16, 30], traits that can be associated with some types of heroes [15] and make it ideal for the "moe-protagonist" archetype [30]. However, in male characters, the color can represent malicious or wicked characters [16], explaining why it beats only gray as an indicator of being a male protagonist.

For male characters, black hair also acted as a good indicator of being a main character, a pattern reflected by the model but not by the participants. This discrepancy is possibly due to black being a very common color for hair [8, 30], second only to brown in our dataset, 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 [16], and despite its dark hue, it usually carries positive connotations, making it ideal for the lone wolf type of character [16], that can play the role of hero in many narratives [34]. 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 participants 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 [16], and these traits can be linked to protagonistic roles [34], possibly explaining, alongside its rarity, why audiences tend to attribute such importance to hair color. However, having white hair can also mean a detachment from reality [30], which may turn antiheroes into villains [34], 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 associated with the color.

Overall, this analysis shows that, when tackling tasks with unclear visual patterns, it is important to consider how models establish the correlation between different visual features and their final results. Such insights can reveal potential biases within the models, which often remain hidden when the underlying patterns are not well understood.

5 CONCLUSIONS

The results of this work show not only how challenging the task of narrative classification can be for humans, but also that deep-learning models can be considerably more effective in learning the intricate patterns that correlate certain visual features and the final classification. It is also clear, however, how the classification process can exaggerate certain biases and overestimate their actual

Table 1: Performance metrics for each classifier. Precision and recall values were calculated as weighted averages of the two classes.

Classifier	Accuracy (%)	Precision (%)	Recall (%)
Survey	58.82	66.04	46.80
ResNet	64.56	66.90	62.74
FM	76.93	79.72	73.93

Table 2: Percentage impact each visual feature had on a character's chance of being labeled as a main character according to each classification method

Feature	Truth (%)	Survey (%)	ResNet (%)	FM (%)
Male	-6.54	-11.23	-11.36	-9.26
Female	5.37	9.36	9.38	8.18
Young	8.81	9.94	10.28	10.03
Old	-61.30	-69.18	-71.46	-67.54
Short Hair	-3.04	-9.15	-5.24	-2.13
Long Hair	5.55	16.87	9.66	4.48
Black Hair	1.50	-13.56	16.22	10.93
Blonde Hair	7.72	28.72	8.93	9.01
Blue Hair	16.83	17.70	12.35	15.49
Brown Hair	-0.43	-12.83	-4.47	-3.21
Gray Hair	-49.71	-55.82	-62.61	-56.46
Green Hair	-12.14	-2.25	1.11	-7.97
Orange Hair	3.50	5.36	8.55	4.53
Pink Hair	15.27	42.84	23.04	21.74
Purple Hair	14.15	8.77	-10.87	2.03
Red Hair	4.84	27.09	28.11	12.36
White Hair	-13.59	16.72	-24.02	-16.41

impact, reaffirming the importance of being aware of these patterns and addressing them whenever possible.

The fact that our classifiers could learn many of the same patterns shown by humans, even if some of those patterns do not reflect reality. Therefore, it means that this kind of model can be used to find the implicit patterns and biases humans use when labeling 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. This, however, must be considered with extreme care when dealing with tasks where such a biased view is undesirable and can be harmful.

When it comes to the specific task we studied, it has been shown that covariance pooling [21] can be used to improve performance in extracting features from anime faces [11] 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.

It could also prove beneficial to expand this kind of study to other tasks in the field of multimedia that share many of its underlying challenges. It can, however, be difficult to assemble the same volume of data for those as we were able to acquire for this work.

REFERENCES

[1] Geoffrey E. Hinton Alex Krizhevsky, Ilya Sutskever. 2012. ImageNet Classification with Deep Convolutional Neural Networks. *NIPS* (2012). https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[2] Joseph Campbell. 2012. *The Hero with a Thousand Faces* (3 ed.). New World Library.

[3] Wenhui Cheng. 2009. Analysis on nationality of Japanese Anime color. *IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design* (nov 2009).

[4] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, Honolulu, HI, USA, 1251–1258. https://doi.org/10.1109/CVPR.2017.195

[5] Joanne Cantor Cynthia Hoffner. 1985. Developmental differences in responses to a television character's appearance and behavior. *Developmental Psychology* 21, 6 (1985).

[6] Joanne Cantor Cynthia Hoffner. 1991. *Responding to the screen: Reception and reaction processes.* Chapter Perceiving and responding to mass media characters, 63–101.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. *IEEE conference on computer vision and pattern recognition* (may 2009). https://doi.org/10.1109

[8] Alex Fox. 2013. *A Viewer's Guide to Anime Hair's Meaning.* Retrieved June 2, 2022 from https://www.forevergeek.com/anime-hairs-meaning/

[9] Bruno Rocha Gomes, Antonio J. G. Busson, José Boaro, and Sérgio Colcher. 2023. Realistic Facial Deep Fakes Detection Through Self-Supervised Features Generated by a Self-Distilled Vision Transformer. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web (WebMedia '23)*. Association for Computing Machinery, New York, NY, USA, 177–183. https://doi.org/10.1145/3617023.3617047

[10] Timothy Hickson. 2019. *On Writing and Worldbuilding*. Vol. 1.

[11] Ivan Jesus, Jessica Cardoso, Antonio Busson, Álan Guedes, Sérgio Colcher, and Ruy Milidiú. 2019. A CNN-based Tool to Index Emotion on Anime Character Stickers. *2019 IEEE International Symposium on Multimedia (ISM)* (december 2019). https://doi.org/10.1109/ISM46123.2019.00071

[12] Ivan Jesus, Jessica Cardoso, Antonio Jose G. Busson, Álan Livio Guedes, Sérgio Colcher, and Ruy Luiz Milidiú. 2019. A CNN-Based Tool to Index Emotion on Anime Character Stickers. In *2019 IEEE International Symposium on Multimedia (ISM)*. 319–3193. https://doi.org/10.1109/ISM46123.2019.00071

[13] Yanghua Jin, Jiakai Zhang, Minjun Li, Yingtao Tian, Huachun Zhu, and Zhihao Fang. 2017. Towards the Automatic Anime Characters Creation with Generative Adversarial Networks. *Comiket 92 (Summer 2017), Tokyo Big Sight* (aug 2017). https://doi.org/1708.05509

[14] Shaoqing Ren Kaiming He, Xiangyu Zhang and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *IEEE conference on computer vision and pattern recognition* (dec 2015). https://doi.org/1512.03385

[15] Michael Weber Katja Rogers, Maria Aufheimer and Lennart E. Nacke. 2018. Towards the Visual Design of Non-Player Characters for Narrative Roles. *Graphics Interface Conference* (may 2018). https://graphicsinterface.org/wp-content/uploads/gi2018-21.pdf

[16] Kevin. 2017. *Anime Hair - Colors and Hairstyles and their Meanings.* Retrieved June 2, 2022 from https://skdesu.com/en/hair-in-anime-colors-and-hairstyles-and-their-meanings/

[17] Hayeon Kim, Eun-Cheol Lee, Yongseok Seo, Dong-Hyuck Im, and In-Kwon Lee. 2021. Character Detection in Animated Movies Using Multi-Style Adaptation and Visual Attention. *IEEE Transactions on Multimedia* 23 (2021), 1990–2004. https://doi.org/10.1109/TMM.2020.3006372

[18] Kang-Ming Chang Kun Liu, Jun-Hong Chen. 2019. A Study of Facial Features of American and Japanese Cartoon Characters. *Symmetry* (may 2019). https://www.mdpi.com/2073-8994/11/5/664/htm

[19] Jochen Laubrock and David Dubray. 2018. CNN-Based classification of illustrator style in graphic novels: which features contribute most?. In *International Conference on Multimedia Modeling*. Springer, 684–695.

[20] Kun Liu, Kang-Ming Chang, Ying-Ju Liu, and Jun-Hong Chen. 2020. Animated Character Style Investigation with Decision Tree Classification. *Symmetry* (jul 2020). https://www.mdpi.com/2073-8994/12/8/1261/pdf

[21] Mengyi Liu, Ruiping Wang, Shaoxin Li, S. Shan, Zhiwu Huang, and Xilin Chen. 2014. Combining Multiple Kernel Methods on Riemannian Manifold for Emotion Recognition in the Wild. *Proceedings of the 16th International Conference on Multimodal Interaction* (2014).

[22] Yusuke Matsui Masaki Saito. 2015. Illustration2Vec: A Semantic Vector Representation of Illustrations. *SIGGRAPH Asia 2015 Technical Briefs* (2015). https://www.gwern.net/docs/anime/2015-saito.pdf

[23] Paulo Renato C. Mendes, Antonio José G. Busson, Sérgio Colcher, Daniel Schwabe, Álan Lívio V. Guedes, and Carlos Laufer. 2020. A Cluster-Matching-Based Method

for Video Face Recognition. In *Proceedings of the Brazilian Symposium on Multimedia and the Web* (São Luís, Brazil) *(WebMedia '20)*. Association for Computing Machinery, New York, NY, USA, 97–104. https://doi.org/10.1145/3428658.3430967

[24] Ping Wang Minghua Liu. 2010. Study on Image Design in Animation. *Asian Social Science* 6, 4 (apr 2010). https://pdfs.semanticscholar.org/6202/af33ab7918cafa3622010e62d245f12d0713.pdf

[25] Walter Mischel Nancy Cantor. 1979. *Advances in experimental social psychology.* Elsevier Inc. Prototypes in person perception.

[26] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 2024. DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research* (2024). https://openreview.net/forum?id=a68SUt6zFt Featured Certification.

[27] Gerald Prince. 1989. *A dictionary of narratology* (1 ed.). University of Nebraska Press.

[28] Vladimir Propp. 1968. *Morphology of the Folktale* (1 ed.). University of Texas Press.

[29] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. (feb 2021). https://doi.org/2103.00020

[30] Martin Schneider. 2015. *Anime Hair Colors: Do They Carry Any Significant Meaning In Japanese Culture?* Retrieved June 2, 2022 from https://thoughtcatalog.com/martin-schneider/2015/02/anime-hair-colors-do-they-carry-any-significant-meaning-in-japanese-culture/

[31] Frederik L. Schodt. 1996. *Dreamland Japan Writings On Modern Manga.* Stone Bridge Press.

[32] Anders Tychsen, Michael Hitchens, and Thea Brolund. 2008. Character play: the use of game characters in multi-player role-playing games across platforms. *Comput. Entertain.* 6, 2, Article 22 (July 2008), 24 pages. https://doi.org/10.1145/1371216.1371225

[33] Valentin. 2013. *Why do Anime Characters have Big Eyes?* Retrieved June 2, 2022 from 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

[34] Christopher Vogler. 1998. *The Writer's Journey* (3 ed.). Michael Wiese Productions.