

Ablation Study of Domain-Specific Focal Layer for Galaxy Image Classification with Convolutional Neural Networks

Tomas MŪŽAS, Andrius Vytautas MISIUKAS MISIŪNAS, Tadas MEŠKAUSKAS

Vilnius University, Faculty of Mathematics and Informatics, Institute of Computer Science
Didlaukio 47, Vilnius LT-08303, Lithuania

`tomas.muzas@mif.vu.lt`, `andrius.misiukas@mif.vu.lt`,
`tadas.meskauskas@mif.vu.lt`

ORCID 0000-0002-2346-724X, ORCID 0000-0002-9279-6194, ORCID 0000-0002-2169-9135

Abstract. The aim of this research was to perform an ablation study of the domain-specific Focal convolution layer, proposed in our previous study. The study focused primarily on the images from ground telescopes – Sloan Foundation and DECam, though Hubble space telescope data was explored as well. The applicability of the Focal layer was studied in two aspects. The first one was to verify whether a regular convolution can be replaced with the Focal to improve results on existing architectures. It was evaluated using two architectures of other authors, each having two variations: Original and Focal (all convolutions are Focal). The second aspect was to evaluate the impact of the Focal layer in isolation on our two best architectures from the previous study, each having three variations: NoFocal (all convolutions are regular), Original, and Focal. For ground-based data, the Focal variations of both other author architectures have achieved a statistically significant improvement of F1 scores compared to the Original variations. However, for both of our architectures, the best results were demonstrated by the Original variations – a mix of both regular and Focal layers – indicating that the layer should be applied conservatively. Additionally, based on the Original variations of both of our architectures, the layer has the highest impact when placed in the initial layers of the network, whereas its application in deeper layers should be avoided. As for the Hubble space telescope data, the Focal layer impact was minimal, what may have been caused by the absence of atmospheric turbulence or background objects.

Keywords: Galaxy morphology classification, Galaxy Zoo, Ablation Study, Convolutional neural networks, Domain-specific Layers

1 Introduction

Recently, the galaxy classification process has become more automated by using convolutional neural networks (hereinafter CNN), trained on the data of volunteer-based

classification projects, such as Galaxy Zoo (Lintott et al., 2008) and its successors. While some impressive spiral vs elliptical results have already been achieved on the brightest galaxies, such as the 99.5% accuracy by Barchi et al. (2020) (see Section 2 for more details), galaxy classification still poses challenges – the classification accuracy decreases when using dimmer galaxies, as well as when more detailed classes are introduced (Barchi et al., 2020).

Besides the traditional means of improving classification results by using augmentations or extracting generic features that describe morphology, an alternative is to utilize domain-specific observations of the galaxy images. For example, in our previous study (Mūžas et al., 2026), we have introduced a Focal convolution layer (see Section 2 for details) intended to help models focus on the galaxy in the center of the image. The new models that contained this layer outperformed the models of other authors when trained without augmentations. However, during that tuning process, multiple different hyperparameters were allowed to change, and the effect of the Focal layer was not fully evaluated in isolation.

Therefore, this study aims to fill this gap by conducting an ablation study of the Focal layer. The ablation is performed in two aspects. The aim of the first aspect is to verify whether the regular convolution layers can simply be replaced with the Focal layers to achieve better results. For this purpose, two CNN architectures of other authors were used – Cavanagh et al. (2021) and Cheng et al. (2020), each having two variations – Original (no modifications, all convolutions are regular), and Focal (all convolutions are Focal). The second aspect was to evaluate the effect of the Focal layer in isolation on the two best architectures from our previous study (Mūžas et al., 2026) – *Top10Loss*, and *BestLoss*. Since both of those architectures originally contained a combination of regular and Focal convolutions, three variations were tested for this aspect – NoFocal (all convolutions are regular), Original (combination of regular and Focal convolutions), and Focal (all convolutions are Focal).

Since the studies of Cavanagh et al. (2021), Cheng et al. (2020), and our previous study Mūžas et al. (2026) used different datasets and number of classes, it was important to ensure a standardized environment to evaluate the applicability of the Focal layer. For this purpose, all aforementioned architectures and their variations were modified to solve a binary (spiral vs. elliptical) problem, and were evaluated using the same datasets, consisting of reliably classified galaxies (see Section 3.1). Primarily, two large datasets were used that covered data from ground telescopes – the Sloan Foundation telescope (Gunn et al., 2006) (referred to as SDSS due to the survey name), and the Dark Energy Camera (Flaugher et al., 2015) (referred to as DECam). It was hypothesized that the Focal layer would have the most significant effect on the ground-based data, as the images are affected by the Earth’s atmosphere or may contain more background objects (stars or other galaxies), which may not necessarily be present in space images. Nonetheless, for the completeness of the study, experiments were also carried out with an additional dataset from the Hubble space telescope (Jakobsen and Laurance, 1989).

The ablation study has shown that the Focal layer helps improve the F1 score when classifying images from ground-based telescopes (see Sections 5.2 and 5.3), however, it should be applied sparingly and in the initial layers of the architectures. In the case of the Hubble space telescope data, the effect of the Focal layer was minimal (see Section

5.4). However, to fully confirm the hypothesis that atmospheric turbulence or background objects were limiting factors requires further investigation.

The paper is structured as follows: in Section 2, related work is discussed. Section 3 describes the data gathering and filtering process. Experiment setup is discussed in Section 4. The in-depth analysis and discussion of the results are presented in Section 5. Finally, conclusions and future work ideas are outlined in Section 6.

2 Related Work

CNNs have already been successfully applied in the field of galaxy classification. For example, using the brightest galaxies, Barchi et al. (2020) have achieved near-perfect results of 99.5% when classifying galaxies into spiral vs. elliptical. Similarly, Cheng et al. (2020) have used their architecture to achieve 99.4% accuracy, while Cavanagh et al. (2021) achieved 98% accuracy. As for more classes, Mittal et al. (2020) have achieved 98% accuracy on a small subset of Galaxy Zoo data, while Zhu et al. (2019) have achieved up to 95.2% accuracy on 5-class data.

However, as shown by the same study of Barchi et al. (2020), the accuracy drops when dimmer galaxies are introduced – CNNs then achieve 98.7%. Additionally, when classification into more classes is performed, the accuracy drops significantly – extending the problem to 9 classes results in 75.7% accuracy for the brightest galaxies, while using 11 classes reduces accuracy to 65.2%.

One of the simplest approaches to improving galaxy classification accuracy is using image augmentation techniques, such as image flipping (Cavanagh et al., 2021; Dieleman et al., 2015; Mittal et al., 2020), rotation (Cavanagh et al., 2021; Cheng et al., 2020; Dieleman et al., 2015) or zoom (Dieleman et al., 2015; Khalifa et al., 2018; Zhu et al., 2019). These techniques allow to create additional training data without changing the galaxy class represented in the image.

Another approach is extracting the generic features that help identify the morphology of the galaxy. For example, Wei et al. (2023) used Deformable Convolution Networks to capture the internal structure of the galaxy, while Li et al. (2023) and Katebi et al. (2019) have explored Capsule Networks to obtain rotation-invariant, spatially-aware features.

The third option would be to utilize the domain-specific observations about the galaxy images and their morphology. This option was explored in our previous study (Mūžas et al., 2026), where we have proposed two domain-specific layers – Focal and Wave – that take galaxy image specifics into account. To the best of our knowledge, domain-specific approach has not yet been thoroughly studied, with the closest approach being the soft attention layer used by Gupta et al. (2024), which dynamically focused model attention to multiple regions of the image.

In our study (Mūžas et al., 2026), the most significant impact to the F1 score was shown by the Focal layer. It was designed to dim the background objects, such as stars or other galaxies, to help model focus on the main galaxy in the center of the image. The dimming is controlled via the parameter *factor*, whose larger values signify faster dimming. Its effect is illustrated in Figure 1.

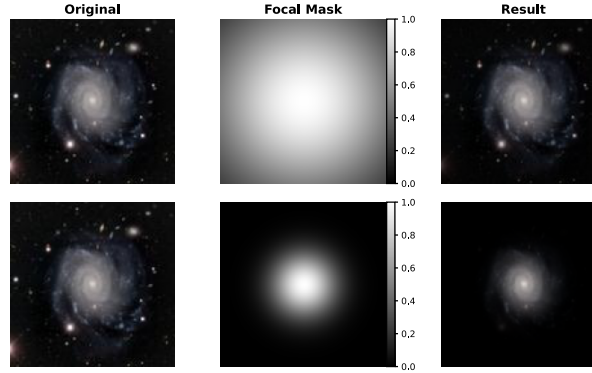


Fig. 1: The illustration of the effects of the Focal layer with different $factor$ values. The top row represents the effect with $factor = 1$, bottom – with $factor = 10$. In each row, the left image is the original galaxy image, middle – visualization of the Focal layer mask with the corresponding $factor$ value, and right – the resulting image after applying the mask.

During the tuning process, it was often preferred over the regular convolution layers. Models with this layer outperformed the aforementioned Cheng et al. (2020) and Cavanagh et al. (2021) models when trained without augmentations. Additionally, the models achieved similar results both with and without augmentations, implying that the layer already helps with regularization (Mūžas et al., 2026).

However, during the tuning process, multiple different hyperparameters were explored – depth, convolution channels and kernel, dropout, dense depth and channels. (Mūžas et al., 2026). Even though the best models often preferred the Focal layer over convolution layers, some of them even fully consisting of the Focal layer, other parameters may have also contributed to the model improvements.

Therefore, this study aims to conduct an ablation study of the Focal layer, using the Cavanagh et al. (2021) and Cheng et al. (2020) architectures, as well as our two best architectures – *BestLoss* and *Top10Loss* – from our previous study (Section 4 for more details). The ablation study is conducted primarily using the data from ground telescopes, though space telescope data is also discussed (see Sections 3 and 5.4).

3 Data

This section describes the pipeline for creating datasets containing reliable classifications.

3.1 Cross-matching the Datasets

This study mainly focuses on the data from the ground telescopes, as covered by Galaxy Zoo 1 (GZ1) (Lintott et al., 2008), Galaxy Zoo 2 (GZ2) (Willett et al., 2013), and the

Galaxy Zoo DECaLS (GZD) (Walmsley et al., 2021) projects. The data for the GZ1 and GZ2 projects was obtained using the Sloan Foundation ground telescope (Gunn et al., 2006), while the data for GZD was collected with the Dark Energy Camera (Flaugher et al., 2015).

However, for the completeness of the study, the data from the Hubble space telescope (Jakobsen and Laurance, 1989) was studied as well, which was covered by GZ Hubble (GZH) (Willett et al., 2016) and GZ CANDELS (GZC) (Simmons et al., 2016) projects.

An important aspect of the GZ datasets is that they are produced by combining the votes of volunteers, who are not required to have any prior astrophysics background. Even though final classifications were assigned based on the multiple votes (such as an average of 42 votes in GZ2 (Willett et al., 2013)), the classifications may still not be accurate. Additionally, the same galaxy can be observed by multiple telescopes, what may reveal more previously unseen features, hence the classifications may also differ between datasets.

To address the uncertainty of volunteer classifications, a data merging process is applied. The process is described as follows:

1. Determine galaxy class based on the debiased vote fractions (Lintott et al., 2008; Willett et al., 2013):
 - (a) Spiral – galaxy has Spiral flag, and no Elliptical flag.
 - (b) Elliptical – galaxy has Elliptical flag, and no Spiral flag.
 - (c) Uncertain – all remaining options.
 For GZ1 dataset, Spiral and Elliptical flags were assigned based on the author proposed classes. For the GZ2 flags were already provided. For the GZD dataset, a corresponding flag was set if the debiased probability was at least 0.5. In all datasets, except GZ1, artifacts were removed based on the artifact flags.
2. Remove all galaxies with uncertain classification.
3. Group galaxies by the dataset IDs. This step mostly covered galaxies from GZ1 and GZ2 datasets, as their IDs overlap.
4. If galaxy classifications are not the same in all datasets in which it is present, all entries of that galaxy are removed. In other words, such galaxy is considered ambiguous, and is therefore not included in the final dataset.
5. If all classifications throughout datasets agree, leave only a single entry of the galaxy. However, the information regarding which datasets the galaxy was classified in is preserved.
6. Group galaxies by the center coordinates – two entries are considered to be the same galaxy if they are within 0.5 arcsec. This results in the same galaxy having 1 or more classifications.
7. Repeat process from steps 4 and 5.

It is important to note that GZH and GZC project data was included into the filtering process as well, slightly impacting the filtering outcome for overlapping galaxies in case classes did not match.

After this process, SDSS dataset had 221,582 spiral and 83,607 elliptical, while DECam had 98,009 spiral and 17,584 elliptical galaxies. The Hubble space telescope dataset contained a total of 45,439 galaxies.

3.2 Images

GZ1 and GZ2 images were collected using SkyServer’s ImgCutout¹ tool. As for all other datasets, images from the Galaxy Datasets package² were used. Images were cropped from the original size of 424×424 to 224×224 to remove the majority of the background and artifacts, and then resized to a final size of 128×128 to conserve computational resources. The same image size reduction pipeline was also used by Jiménez et al. (2020).

It is important to highlight several aspects. First of all, images for GZ DECaLS campaigns GZD-1 and GZD-2 were not available, only for the GZD-5. Additionally, a single galaxy can be in more than one dataset, meaning that in some cases, more than one image is available for the galaxy.

4 Experiments

As discussed in Section 2, during the tuning process of our previous study (Mūžas et al., 2026), multiple values of hyperparameters were allowed to change. Therefore, before conducting this ablation study, an interim two-step study was conducted to identify the hyperparameters that have the most impact on the results, with one of the hyperparameters being whether to apply the Focal layer. In the first step, 1000 random architectures were explored, while in the second, some hyperparameter values were fixed, and 297 architectures were further explored. In both cases, the Focal layer demonstrated the potential to improve metrics with different hyperparameter combinations. Therefore, it was decided to evaluate the effect of the Focal layer in isolation.

4.1 Architectures and Their Variations

The applicability of the Focal layer was evaluated in two different aspects. The first aspect was to evaluate whether it would be beneficial to simply replace the convolutional layers in existing CNN architectures. For this purpose, two architectures that achieved great results in binary classification were used – Cavanagh et al. (2021) and Cheng et al. (2020). Two variations of each architecture were explored:

1. Original
 - Convolutions: all regular, as described in original papers.
 - Purpose: for baseline results.
2. Focal
 - Convolutions: all Focal, $factor = 10$.
 - Purpose: to evaluate whether replacing regular convolutions with Focal will improve results.

The second aspect was to verify whether the Focal layer had significant impact on the two best architectures from our previous study (Mūžas et al., 2026) – *Top10Loss*, initially designed for 4 classes, and *BestLoss*, designed for 11 class task. Since both architectures are a combination of regular and Focal convolutions, each architecture had 3 variations this time:

¹ <http://skyserverice.pha.jhu.edu/DR7/ImgCutout/getjpeg.aspx>

² <https://pypi.org/project/galaxy-datasets/>

1. NoFocal
 - Convolutions: all regular, as described in (Mūžas et al., 2026).
 - Purpose: to evaluate how the proposed models would perform if no Focal layers were present.
2. Original
 - Convolutions: combination of regular and Focal, as in the original paper. For the *Top10Loss*, the Wave layer was replaced with a regular convolution to avoid introducing other layers.
 - Purpose: to evaluate a combination of regular and Focal convolutions.
3. Focal
 - Convolutions: all Focal, $factor = 10$.
 - Purpose: to evaluate whether performance would have improved if all convolutions were Focal.

The choice of $factor = 10$ in Focal layers represents the most aggressive dimming of the background, while still preserving the galaxy in the center, allowing to evaluate the layer at its full potential. Additional motivation comes from the previous study (Mūžas et al., 2026), where during the tuning process, higher $factor$ values were preferred. Convolution kernel and channel parameters, along with all other layers, remained unchanged throughout the variations.

Additionally, all regular convolution layers, including those of the other authors', will use padding parameter "same" that preserves the output dimensions of the convolution. This was done to replicate the setup of our previous study (Mūžas et al., 2026).

Finally, it is important to note that while the models of Cavanagh et al. (2021) and Cheng et al. (2020) architectures have displayed great results (see Section 2), both authors have used relatively small datasets – Cavanagh et al. (2021) used Nair and Abraham (2010) dataset of 14,304 expert-classified galaxies, while Cheng et al. (2020) used only 2,862 galaxies. Additionally, our *Top10Loss* and *BestLoss* architectures were initially designed for classification into 4 and 11 classes respectively. Hence, since all models were evaluated using different datasets and class configurations, direct comparison with published metrics would be misleading. Therefore, to isolate the effect of the Focal layer, all architectures and their variations were adapted for binary (spiral vs elliptical) classification, and evaluated using the same datasets, comprised of significantly more reliably classified galaxies (see Section 3).

The final implementations of the two variations of the Cavanagh et al. (2021) and Cheng et al. (2020) architectures to verify the first aspect are explained in detail in Tables 1 and 2 respectively. The three variations of our *Top10Loss* and *BestLoss* architectures to evaluate the second aspect are presented in Tables 3 and 4.

4.2 Setup

The training process was conducted using 10-fold cross-validation, using Adam (Kingma and Ba, 2014) optimizer with default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$, and a categorical cross-entropy loss function.

The models were allowed to train for up to 500 epochs, however, the process was stopped early if the validation F1 metric did not improve for 10 consecutive epochs.

Table 1: Original and Focal variations of the Cavanagh et al. (2021) architecture. FC stands for fully connected layer, BN - batch normalization, MP - max pooling 2x2. ReLU is used in regular and Focal convolutions, and FC layers.

Original	Focal
Conv., 7x7, 32 ch.	Focal (10), 7x7, 32 ch.
BN + MP	BN + MP
Conv., 5x5, 64 ch.	Focal (10), 5x5, 64 ch.
Conv., 5x5, 64 ch.	Focal (10), 5x5, 64 ch.
BN + MP	BN + MP
Conv., 3x3, 128 ch.	Focal (10), 3x3, 128 ch.
BN + MP	BN + MP
Flatten	Flatten
Dropout 0.5	Dropout 0.5
FC, 256/256	FC, 256/256
FC, 2 (softmax)	FC, 2 (softmax)

Table 2: The Original and Focal variations of Cheng et al. (2020) architecture. For abbreviations see Table 1

Original	Focal
Conv., 3x3, 32 ch.	Focal (10), 3x3, 32 ch.
MP	MP
Conv., 3x3, 64 ch.	Focal (10), 3x3, 64 ch.
MP	MP
Conv., 2x2, 128 ch.	Focal (10), 2x2, 128 ch.
MP	MP
Flatten	Flatten
Dropout 0.5	Dropout 0.5
FC, 1024/1024	FC, 1024/1024
Dropout 0.5	Dropout 0.5
FC, 2 (softmax)	FC, 2 (softmax)

Results (Section 5) are reported from the epoch where the overall best results were achieved.

The majority of the computation was performed on the supercomputer at Vilnius University, IT Research Center, using the NVIDIA Tesla V100 graphical processing unit (GPU). Some of the computations were performed on the NVIDIA RTX 4080 GPU. As the model training or inference speed was not evaluated in this study, the choice of GPU did not have any effect on the reported results.

Table 3: NoFocal, Original, and Focal variations of the *Top10Loss* architecture (Mūžas et al., 2026). The last Wave layer of the Original variation was replaced with regular convolution. For abbreviations see Table 1.

NoFocal	Original	Focal
Conv, 7x7, 64 ch.	Focal (8), 7x7, 64 ch.	Focal (10), 7x7, 64 ch.
BN + MP	BN + MP	BN + MP
Conv. 5x5, 64 ch.,	Focal (5), 5x5, 64 ch.	Focal (10), 5x5, 64 ch.
BN + MP	BN + MP	BN + MP
Conv., 7x7, 32 ch.	Conv., 7x7, 32 ch.	Focal (10), 7x7, 32 ch.
BN + MP	BN + MP	BN + MP
Conv., 5x5, 32 ch.	Conv., 5x5, 32 ch.*	Focal (10), 5x5, 32 ch.
BN + MP	BN + MP	BN + MP
Flatten	Flatten	Flatten
Dropout 0.5	Dropout 0.5	Dropout 0.5
FC, 32/160/160	FC, 32/160/160	FC, 32/160/160
FC, 2, (softmax)	FC, 2, (softmax)	FC, 2, (softmax)

Table 4: NoFocal, Original, and Focal variations of the *BestLoss* architecture (Mūžas et al., 2026). For abbreviations see Table 1.

NoFocal	Original	Focal
Conv, 3x3, 48 ch.	Focal (10), 3x3, 48 ch.	Focal (10), 3x3, 48 ch.
BN + MP	BN + MP	BN + MP
Conv. 7x7, 32 ch.	Conv. 7x7, 32 ch.	Focal (10), 7x7, 32 ch.
BN + MP	BN + MP	BN + MP
Conv., 7x7, 48 ch.	Conv., 7x7, 48 ch.	Focal (10), 7x7, 48 ch.
BN + MP	BN + MP	BN + MP
Conv., 3x3, 32 ch.	Conv., 3x3, 32 ch.	Focal (10), 3x3, 32 ch.
BN + MP	BN + MP	BN + MP
Flatten	Flatten	Flatten
Dropout 0.5	Dropout 0.5	Dropout 0.5
FC, 224/192	FC, 224/192	FC, 224/192
FC, 2, (softmax)	FC, 2, (softmax)	FC, 2, (softmax)

5 Results and Discussion

5.1 Model Comparison

The models in this section are compared based on their F1 score, as the datasets are heavily imbalanced (see Section 3). All values reported in this section are an average of the F1 scores of all 10 folds.

To compare the results of the models across the folds, the Wilcoxon (1945) signed rank test was used, since it does not assume a normal distribution – with only 10 data

points (folds), the Shapiro-Wilk (Shapiro and Wilk, 1965) metric power is low, preventing a reliable verification of the normal distribution hypothesis. The significance threshold of 0.05 was applied for the p-value, below which the results of two models are considered different.

5.2 Applicability on Other Architectures

The results of replacing regular convolution layers with the Focal layer in Cavanagh et al. (2021) and Cheng et al. (2020) architectures, using ground-based datasets are shown in Table 5

Table 5: Results of replacing regular convolution with Focal layer in architectures of other authors. Best results are highlighted in bold.

(a) DECam			
Model	Original F1 (σ)	Focal F1 (σ)	p-value
Cavanagh	80.34 (0.60)	81.56 (0.64)	0.002
Cheng	78.99 (0.73)	81.60 (0.72)	0.002
(b) SDSS			
Model	Original F1 (σ)	Focal F1 (σ)	p-value
Cavanagh	93.56 (0.12)	94.07 (0.15)	0.002
Cheng	93.44 (0.17)	94.14 (0.15)	0.002

Based on the Table 5, it can be observed that applying the Focal layer on other authors' architectures improves the F1 score by a statistically significant amount – for both SDSS and DECam datasets, and for both Cavanagh and Cheng models, the p-values were 0.002.

These results show that the classification results can be improved simply by replacing regular convolutions with the Focal. However, it is important to note that both architectures are linear (there are no branches). Additionally, in both cases, due to the "same" padding parameter, the convolution outputs in all layers preserve large output dimensions (only each max pooling operation decreases output size by 2). Therefore, the same effect may not persist if models use branches or use convolutions with small output dimensions.

5.3 Effect on Our Architectures

The results of all 3 variations of our *Top10Loss* and *BestLoss* architectures are shown in Table 6.

Based on the Table 6, the results show that the F1 metric is not always improved when using the Focal variation. For the DECam data, Focal variations of both *Top10Loss*

Table 6: Results of replacing regular convolution with Focal layer in our architectures from previous study. Both p-values are calculated against the baseline model. Best results are highlighted in bold.

(a) DECam					
Model	NoFocal F1 (σ)	Original F1 (σ)	p-value	Focal F1 (σ)	p-value
<i>BestLoss</i>	81.57 (0.29)	82.23 (0.62)	0.006	81.38 (0.54)	0.232
<i>Top10Loss</i>	81.41 (0.74)	82.07 (0.69)	0.010	80.98 (0.95)	0.131
(b) SDSS					
Model	NoFocal F1 (σ)	Original F1 (σ)	p-value	Focal F1 (σ)	p-value
<i>BestLoss</i>	94.13 (0.15)	94.15 (0.10)	0.770	93.99 (0.16)	0.006
<i>Top10Loss</i>	94.09 (0.14)	94.20 (0.18)	0.160	94.01 (0.19)	0.193

and *BestLoss* architectures performed similarly to the NoFocal variations, with p-values of 0.131 and 0.232 respectively. For the SDSS dataset, *Top10Loss* Focal variation performed similarly, though the *BestLoss* has actually produced worse results than the NoFocal variations. This result demonstrates that in some cases, simply replacing all convolutions with Focal may not be beneficial, or sometimes even detrimental.

Nonetheless, it is important to highlight that for both *Top10Loss* and *BestLoss* models and both SDSS and DECam datasets, the best results were shown by the Original variation – a combination of regular and Focal convolutions. In case of DECam data, the improvement of Original variation compared to the NoFocal was significant – *BestLoss* model achieved p-value of 0.006, while *Top10Loss* improved with p-value of 0.01. In case of SDSS data, Original variations of both architectures performed similarly to the NoFocal ones, however, the average F1 values of Original variations were slightly higher.

It is important to highlight that in both *Top10Loss* and *BestLoss* architectures, Focal layers are more effective in the initial feature extraction stages (see Tables 3 and 4), where the background noise is more dominant. In deeper layers, the Focal layer becomes less beneficial. Therefore, it is recommended to use this layer primarily in the initial layers of the network.

5.4 Results of Space Telescope Data

The ground telescopes were chosen as the main focus since the ground telescopes can produce significantly more data, as they are cheaper and easier to operate than the space telescopes. This difference is likely to persist in the future as well due to the high costs of launching and operating space telescopes.

The specifics of the space telescopes are also hypothesized to limit the effectiveness of the Focal layer. One reason could be that space-based images are not influenced by Earth's atmospheric turbulence, whereas in ground-based images, the focal layer counteracts these effects by adjusting the focus on the galaxy. Another limiting factor could

be that the Focal layer is most effective when other background objects are present, yet since the space telescopes usually observe more distant galaxies, the radial resolution may include fewer background objects.

However, for the sake of completeness, the same two aspects described in Sections 5.2 and 5.3 were carried out on the data of the Hubble space telescope (see Section 3) to verify the hypothesis that the Focal layer effect will be limited when no atmospheric turbulence is present in the images.

In both aspects, the models did not demonstrate a statistically significant improvement. When applying the Focal layer to the Cavanagh et al. (2021) and Cheng et al. (2020) architectures, the results of Focal variations were similar to the Original variations. Additionally, neither the Original nor the Focal variations of *BestLoss* and *Top10Loss* models outperformed the NoFocal variant. However, the same observation as for SDSS data persisted that the *BestLoss* Focal variation performed worse than the NoFocal.

Nonetheless, additional study is required to verify that the presence of the atmospheric turbulence or other background objects is the reason behind the limited impact of the layer. Additionally, it would be beneficial to explore different *factor* values, as potentially, lower values could be useful when image quality is better. Finally, different domain-specific layers could be required to improve classification metrics for the space telescope datasets.

6 Conclusions

After conducting an ablation study of the domain-specific Focal layer using both ground and space telescope datasets, it can be concluded that:

1. The Focal layer proves to be the most useful for the ground-based telescope data – SDSS and DECam. It helps improve the F1 metrics, however, it should also be applied conservatively. Based on Table 5, both for Cavanagh et al. (2021) and Cheng et al. (2020), the Focal variations outperformed the Original ones by a statistically significant amount (Wilcoxon signed rank test p-values were all 0.002). However, based on Table 6, for our previous *Top10Loss* and *BestLoss* architectures, the best results were demonstrated by the Original variations, which are a combination of regular and Focal convolutions. For DECam data, the Original variations achieved better results than the NoFocal ones with p-values of 0.006 for *BestLoss*, and 0.01 for *Top10Loss*, while the results for SDSS data were similar. The Focal variations achieved similar, or in case of *BestLoss* and SDSS data, degraded results (p-value 0.006) compared to the NoFocal variations.
2. Based on the results in Table 6, and the structure of our architectures described in Tables 3 and 4, the best results are achieved when the Focal layer is placed in the initial layers of the architecture, and followed by regular convolution in deeper layers.
3. As described in Section 5.4, the impact of the Focal layer is minimal on the data from the Hubble space telescope. The reason behind the limited effect might be due to the absence of atmospheric turbulence or background objects. However, this hypothesis needs to be verified separately.

As for future research, it would be beneficial to explore the effect of different *factor* values for both ground and space-based telescope datasets. Additionally, as discussed in Section 5.4, different domain-specific layers could be designed that are better suited to the specifics of the space-based data.

Acknowledgements

The computational resources for this study were provided by the Information Technology Research Center of Vilnius University. The dataset storage was funded by the research credits provided by Google LLC, number 440629809.

References

- Barchi, P., de Carvalho, R., Rosa, R. et al. (2020). Machine and Deep Learning applied to galaxy morphology - A comparative study, *Astronomy and Computing* **30**.
- Cavanagh, M. K., Bekki, K., Groves, B. A. (2021). Morphological classification of galaxies with deep learning: comparing 3-way and 4-way CNNs, *Monthly Notices of the Royal Astronomical Society* **506**(1), 659–676.
- Cheng, T., Conselice, C. J., Aragón-Salamanca, A. et al. (2020). Optimizing automatic morphological classification of galaxies with machine learning and deep learning using Dark Energy Survey imaging, *Monthly Notices of the Royal Astronomical Society* **493**, 4209–4228.
- Dieleman, S., Willett, K. W., Dambre, J. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction, *Monthly Notices of the Royal Astronomical Society* **450**(2), 1441–1459.
- Flaugher, B., Diehl, H. T., Honscheid, K. et al. (2015). The Dark Energy Camera, *The Astronomical Journal* **150**(5).
- Gunn, J. E., Siegmund, W. A., Mannery, E. J. et al. (2006). The 2.5 m Telescope of the Sloan Digital Sky Survey, *The Astronomical Journal* **131**(4).
- Gupta, A., Kaur, K., Jindal, N. (2024). Predicting galaxy morphology using attention-enhanced resnets, *Earth Science Informatics* pp. 1–12.
- Jakobsen, R., Laurance, R. J. (1989). The Hubble Space Telescope., *ESA Bulletin* **58**, 91–100.
- Jiménez, M., Torres Torres, M., John, R., Triguero, I. (2020). Galaxy Image Classification Based on Citizen Science Data: A Comparative Study, *IEEE Access* **8**, 47232–47246.
- Katebi, R., Zhou, Y., Chornock, R. et al. (2019). Galaxy morphology prediction using capsule networks, *Monthly Notices of the Royal Astronomical Society* **486**(2), 1539–1547.
- Khalifa, N. E., Hamed Taha, M., Hassanien, A. E. et al. (2018). Deep Galaxy V2: Robust Deep Convolutional Neural Networks for Galaxy Morphology Classifications, *2018 International Conference on Computing Sciences and Engineering (ICCSE)*, IEEE, New Jersey, United States, pp. 1–6.
- Kingma, D., Ba, J. (2014). Adam: A Method for Stochastic Optimization, *International Conference on Learning Representations*.
- Li, G., Xu, T., Li, L. et al. (2023). Galaxy morphology classification using multiscale convolution capsule network, *Monthly Notices of the Royal Astronomical Society* **523**(1), 488–497.
- Lintott, C. J., Schawinski, K., Slosar, A. et al. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, *Monthly Notices of the Royal Astronomical Society* **389**(3), 1179–1189.

- Mittal, A., Soorya, A., Nagrath, P. et al. (2020). Data augmentation based morphological classification of galaxies using deep convolutional neural network, *Earth Science Informatics* **13**, 601–617.
- Mūžas, T., Misiukas Misiūnas, A. V., Meškauskas, T. (2026). Evaluation of domain-specific convolutional neural network layers for galaxy classification tasks, *Nonlinear Analysis: Modelling and Control* **31**, 1–18.
- Nair, P. B., Abraham, R. G. (2010). A catalog of detailed visual morphological classifications for 14,034 galaxies in the sloan digital sky survey, *The Astrophysical Journal Supplement Series* **186**(2), 427–456.
- Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples), *Biometrika* **52**(3/4), 591–611.
<http://www.jstor.org/stable/2333709>
- Simmons, B. D., Lintott, C., Willett, K. W. et al. (2016). Galaxy Zoo: quantitative visual morphological classifications for 48 000 galaxies from CANDELS , *Monthly Notices of the Royal Astronomical Society* **464**(4), 4420–4447.
- Walmsley, M., Lintott, C., Geron, T. et al. (2021). Galaxy Zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies, *Monthly Notices of the Royal Astronomical Society* **509**(3), 3966–3988.
- Wei, S., Lu, W., Dai, W. et al. (2023). Galaxy morphological classification of the legacy surveys with deformable convolutional neural networks, *The Astronomical Journal* **167**(1), 29.
<https://dx.doi.org/10.3847/1538-3881/ad10ab>
- Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics Bulletin* **1**(6).
- Willett, K. W., Galloway, M. A., Bamford, S. P. et al. (2016). Galaxy Zoo: morphological classifications for 120 000 galaxies in HST legacy imaging, *Monthly Notices of the Royal Astronomical Society* **464**(4), 4176–4203.
- Willett, K. W., Lintott, C. J., Bamford, S. P. et al. (2013). Galaxy Zoo 2: detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey, *Monthly Notices of the Royal Astronomical Society* **435**(4), 2835–2860.
- Zhu, X., Dai, J., Bian, C. et al. (2019). Galaxy morphology classification with deep convolutional neural networks, *Astrophysics and Space Science* **364**(4).

Received May 23, 2026 , revised September 1, 2026, accepted September 20, 2026