Seasonal Domain Shift in the Global South: Dataset and Deep Features Analysis

Georgios Voulgaris
University of Sussex
gv29@sussex.ac.uk

Andy Philippides
University of Sussex
andrewop@sussex.ac.uk

Jonathan Dolley
University of Sussex
J.Dolley@sussex.ac.uk

Jeremy Reffin
University of Sussex
J.P.Reffin@sussex.ac.uk

Fiona Marshall
University of Sussex
F.Marshall@sussex.ac.uk

Novi Quadrianto
University of Sussex
N.Quadrianto@sussex.ac.uk

Abstract

Domain shifts during seasonal variations are an important aspect affecting the robustness of aerial scene classification and so it is important that such variation is captured within aerial scene datasets. This is more evident in geographic locations in the global South, where aerial coverage is scarcer and the rural and semi-urban landscape varies dramatically between wet and dry seasons. As current datasets do not offer the ability to experiment with domain shifts due to seasonal variations, this work proposes a labelled dataset for classifying land use from aerial images, comprising both wet and dry season data from Ghaziabad in India. Moreover, we conduct a thorough investigation into how image features, namely colour, shape, and texture, influence the accuracy of scene classification. We demonstrate that a combination of an architecture that extracts salient features, with the implementation of a larger receptive field improves classification performance when applied to both shallow or deep architectures by extracting invariant feature representations across domains.

1. Introduction

Classifying land usage in global South is important because sub-urbanization and peri-urbanization are still predominantly occurring in areas that are not subject to land use planning [1, 14]. Because of the availability of satellite images, it would be desirable to perform automatic land use classification from aerial data. However, there is a lack of validated datasets covering land-use in the global South and in particular which contain the large changes to images that are seen between seasons in this region.

The lack of data is an issue because the exceptional performance of Deep Learning models' image classification is in part a result of the availability of labelled training datasets. Because the early large-scale datasets such as ImageNet [12], focused on objects, often in an indoor setting, the majority of models, were trained and tested on objects rather than scenes. However, compared to objects, scene classification is complex as scenes can be comprised of a mixture of natural and man-made objects, which differ in terms of texture and shape. Moreover, when classifying land use, overlapping classes within scenes such as urban green spaces, forest & orchard and agriculture add to the complexity of the classification task. Finally, especially, for aerial scenes, appearances are affected by do-

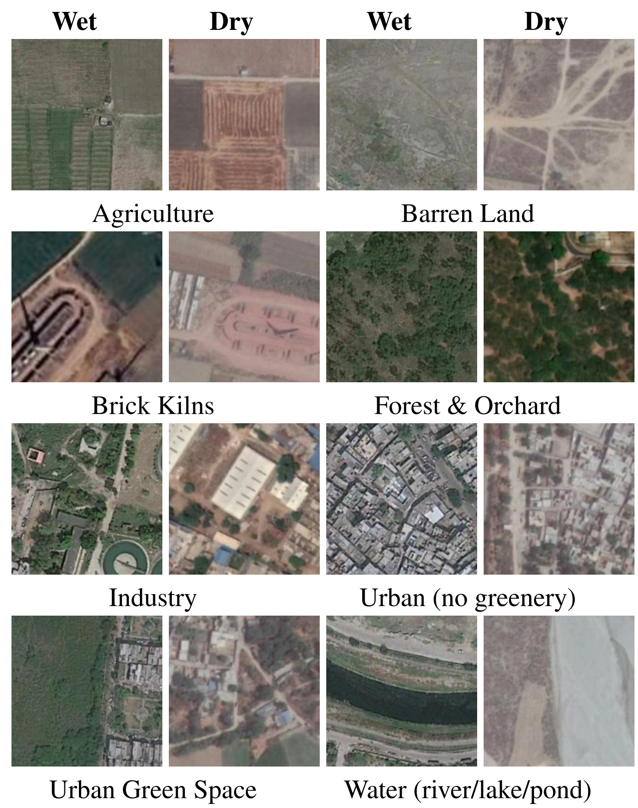

Figure 1. Wet and dry season aerial scenes from Ghaziabad India.

main shifts during seasonal variation which is particularly evident in the global South where rural and semi-urban scenery varies dramatically between wet and dry seasons. Here we therefore present and test a dataset for classifying land use from aerial images, comprising both wet and dry season data from Ghaziabad in India. **Dataset is available at [DOI 10.25377/sussex.22221109]**. We also propose an architecture that improves the performance of both shallow and deep networks in aerial image classification in the presence of domain shifts by extracting salient features through Gabor feature priors and the implementation of a large receptive field.

State-of-the-art aerial datasets While there are a number of datasets of aerial images, they either lack coverage of ecosystems (known as natural capital) which provide for human well-being and quality of life in the global South or do not capture seasonal variations. Earlier datasets depicting aerial scenes such as [24], [25], and [26] were of limited size, comprised of 2,800, 2,100, and 950 aerial images respectively. In recent years, the evolution of remote sensors and unmanned aerial vehicles (UAV), enable the collection of more aerial images, thus allowing the creation of larger, more robust datasets. The work of Xia et al. [23] introduces a large dataset – Aerial Image Data set (AID) – for the task of aerial scene classification. It is comprised of 10,000 images of aerial scenes covering 30 aerial scene types such as airport, bare land, baseball field, and beach. However, AID focuses on global north countries and does not focus on temporal changes.

To tackle the domain generalisation problem across time, whilst improving sub-population performance across regions, Christie et al. [5] propose a Functional Map of the World (fMoW) dataset that combines RGB images (224×224 pixels) and has labels of 62 classes depicting buildings or land use categories and metadata such as the year that the image was acquired and its geographical region. While very large, fMoW covers only a subset of ecosystems and, being satellite images, has a lower spatial resolution than aerial images. Nikolov et al. [15], propose another dataset that varies over time, the Long-term Thermal Drift (LTD) dataset. This dataset comprises of thermal surveillance images from a single location across 8 months combined with metadata that include information such as weather, the day/night cycle, and scene activity. While LTD has a very high temporal resolution and can be used to assess robustness to different domain shifts the images were only taken at the harbor front in Aalborg, Denmark.

In terms of datasets from the global South, Wang et al. [20] have produced the LoveDA dataset to counter their perception that current land-cover datasets lack model transferability because they are biased towards semantic representations. The LoveDA dataset is comprised of 5,987 HSR images with 166,768 annotated objects from three different cities in China: Nanjing, Changzhou, and Wuhan. While this dataset has images of both urban and rural domains, LoveDA focuses on semantic segmentation task and does not capture seasonal variations.

There is therefore a lack of aerial land use datasets which focus on domain shifts due to seasonal variations. A notable exception is Wang et al. [21], which proposes a data set that covers various points of view of Toronto, ranging from airplanes, drones, to cars. The images capture span is 4 years (2009, 2011, 2012, 2013) and depict several seasons. This is close to our proposed dataset, in terms that include data of various seasons from the same area. However, the data are not separated by season and so do not allow easy experimentation against domain shifts. Furthermore, the dataset only focuses on urban areas in Toronto, whilst ours covers remote areas in the global South.

Texture-biased models. Part of deep learning's success in image classification is its ability to automatically extract predictive features from complex, high-dimensional data. These features enable the classifier to generalise from a particular task to other similar ones. The model's generalisation is however dependent on the extracted features in the presence of distribution shift, which can lead to spurious correlations. Spurious correlations are associations that are present in the training data, but not valid for the actual task. For example, if a classifier is trained using wet season data, it can associate Agriculture or Barren Land classes with certain colour features (figure 1) that are dominant in the wet season (i.e. green colour due to flora). In this case, the model will not be able to generalise well when predicting images depicting scenes during the dry season where the flora is scarce (colour brown).

One aim of this work is therefore to use feature priors to control the sensitivity of the training process to spurious correlations. According to Fukushima [7], CNNs are biased towards learning a hierarchy of localized features. Feature priors are useful because they can bias a model towards a particular set of features and significantly impact the architecture's reliance on them and therefore its generalisation ability. For instance, according to Geirhos et al. [8], CNNs trained on ImageNet tend to classify images based on texture features, instead of colour or shape and so are considered to be texture biased. In contrast, classifiers that rely mostly on shape features are considered shape biased.

Herman et al. [11] investigate the argument that CNN architectures are biased towards texture features further, and argue that this is not because of the actual network architectures' inductive biases, but due to the ImageNet training data. Hence, by careful data selection, CNNs can learn to classify based on shape features as well. Here we investigate this point in the context of image scene analysis by training on data from either the wet or dry season only and then testing on the other. As the spurious attributes in both

Figure 2. (a) Original, (b) grey-scale, (c) Canny Edge, (d) Sobel

wet and dry season datasets are colour and texture, we can investigate how the dataset can lead to apparent feature bias and how feature priors can ameliorate this issue.

Contributions The contributions of this work are as follows: 1) A dataset comprised of aerial images of mixed and separated wet/dry season data. 2) Assessment of classifier feature bias when trained on images containing scenes from aerial images. 3) Demonstration that applying a feature prior to the learning process enables better model generalisation in the presence of domain shift, highlighting the importance of appropriate feature priors. 4) Proposal of an architecture which extracts features that are invariant between domains by using a Gabor feature prior in combination with a wide receptive field and demonstration of improved accuracy with a shallow architecture in the presence of domain shift. 5) Implementation of the proposed solution on a deeper architecture, with improved performance on our dataset and benchmark satellite and domain shift data sets.

2. Methods

The purpose of this work is to examine model robustness during data domain shifts in terms of variations between seasons, as well as how applying feature priors to the learning process biases the model to features that allow better model generalisation. To do so, first, the models are trained and evaluated using a dataset that is comprised of mixed-season images. We then examine how the models perform in the presence of domain shifts by training on dry and testing on wet season data and vice versa. In both scenarios, we examine model reliance on the following features: colour, shape, and texture. These features were chosen, because as seen in figure 1, the flora changes the landscape dramatically between seasons and, visually, the features that vary the most are colour, texture and shape. We first describe the datasets and then the models.

2.1. Data

The dataset is comprised of aerial colour images of 256 x 256-pixel size and depicts scenes from nine classes of Ghaziabad India (table 1). The images are of both wet and dry seasons. The aerial imagery was downloaded from Google Earth Pro as maximum resolution images in a grid of 121 rectangles (5000m by 3500m at surface level) covering the Ghaziabad sub-district with an extent of 8593000, 3328500, 8648000, 3367000. Using the historical imagery

Table 1. Mixed, Wet, Dry Season Dataset

Class	Wet	Dry	Mixed
Agriculture	3,278	1,772	5,621
Barren Land	632	301	924
Brick Kilns	968	292	1,008
Forest	936	713	1,516
Orchard Industry	711	726	1,144
Roads	-	-	565
Urban	1,241	566	1,420
Urban Grn Space	584	306	838
Water	404	662	141

selector, data for two time periods were chosen (2013-2014 and 2018-2019). For each time period, dry season (March-May) imagery was selected based on manual visual assessment of clarity, colour tone and cloud cover. The aim was to obtain imagery with high clarity and consistency of colour tone as a priority and to minimise cloud cover and cloud shadow. When dry season imagery was poor quality or obscured by clouds, wet season (June-February) imagery closest to the time period of adjacent tiles (usually January or February) was selected with a focus on getting the highest clarity of imagery. The capture date of each image was recorded in the file names and geo-referencing information was recorded in a metadata file. These images were then uploaded to the https://landscapes.wearepal.ai/ for processing into square tiles and re-sampled at multiple zoom levels. No other image processing or corrections were carried out. Once uploaded to the web-app, the training data was created by using a specially designed interface in the web-app to label tiles by hand for each land classification based on local knowledge of the area.

Colour-biased models. To eliminate colour information in the data, images are pre-processed to grey-scale. Thus, any colour information is lost. Figure 2 (a) shows the original image and 2 (b) the grey-scale one.

Shape-bias models. To eliminate texture information in the data, images are pre-processed by applying edge detection algorithms. For the purpose of this work, two edge detection algorithms were employed: Canny edge detection [6], which creates a binary mask of edges (figure 2 (c)), and Sobel algorithm [17] that results in edge detector that retains a small amount of texture information (figure 2 (d)).

2.2. Models Under Review

2.2.1 CNN5

The first model under consideration is comprised of five hidden convolutional layers. Each convolutional layer includes a convolutional layer, followed by an activation function (ReLU), a maximum pooling layer and a batch normalisation. Furthermore, the last convolutional layer is connected to a global average pooling layer. Global pooling was first mentioned in the work of Lin et al. [13] and became mainstream with ResNet by He et al. [9].

2.2.2 Dilated Gabor Mix Pool CNN (DGCNN)

Dilated convolution was first introduced by Chen et al. [2] [3] as a way of increasing the receptive field for the task of semantic segmentation. According to Wei et al. [22], convolutional kernel receptive fields are enlarged when employing varying dilation rates, which results in transferring the surrounding discriminative information to the discriminative scene regions. In our previous work [19], we showed that a Gabor convolutional layer combined with a mixture of maximum and average pooling improved performance by focusing the classifier on salient features, even in the presence of domain shifts due to seasonal variations. In this paper, we therefore wanted to see if adding dilated convolution could further boost performance.

Specifically, for the Gabor-implemented convolutional layer we used:

$$g(x,y,\omega,\theta,\psi,\sigma) = \exp(-\frac{x'^2+y'^2}{2\sigma^2})\cos(\omega x'+\psi) \quad (1)$$
$$x' = x\cos\theta + y\sin\theta; \quad y' = -x\sin\theta + y\cos\theta,$$

where (x, y) stands for the pixel spatial domain position, θ the filter orientation, σ the standard deviation, and ω the frequency. We use a bank of Gabor filters with frequencies ω_n and orientations θ_m:

$$\omega_n = \frac{\pi}{2}\sqrt{2}^{-(n-1)} \quad \text{where}, n = 1, 2, ..., 5 \quad (2)$$

$$\theta_m = \frac{\pi}{8}(m-1) \quad \text{where}, m = 1, 2, ..., 8. \quad (3)$$

Gabor layer weights were initialised by setting the standard deviation to $\sigma \approx \frac{\pi}{\omega}$. Furthermore, ψ is set by uniform distribution Unif.$(0, \pi)$. Note that the Gabor function parameters are updated during backpropagation.

In addition, due to the information complexity in satellite images, a method that combines maximum and average pooling was applied. Hence on the lowest layer, Maximum pooling was applied, and as incremented to higher layers, it was decreased by 0.2:

$$f_{\text{mix}}(x) = \alpha_l \cdot f_{\max}(x) + (1-\alpha_l) \cdot f_{\text{avg}}(x), \quad (4)$$

where scalar mixing portion $\alpha_l \in [0, 1]$ indicates the max and average combination per layer l.

As an evolution of our previous architecture, similar to [22], a dilated convolutional layer was added between the last convolutional layer and the global average pooling. Specifically, convolutional blocks with multiple dilated rates (i.e. d = 1, 3, 6, 9) were appended to the final convolutional layer, thus localising scene-related regions observed by different receptive fields. Using high dilation rates (i.e. d = 9) can cause inaccuracies by mistakenly highlighting scene-irrelevant regions. To avoid such scenarios, we used equation 5, where the average over the localization maps H_i (i.e i = 3, 6, 9) generated by different dilated convolutional blocks was summed to the localisation map H_0 of the convolutional block with dilation d = 1.

$$H = H_o + \frac{1}{n_d}\sum_{n=1}^{n_d} H_i \quad (5)$$

2.2.3 Dilated Gabor Mix Pool ResNet18 (DilGabMPResNet18)

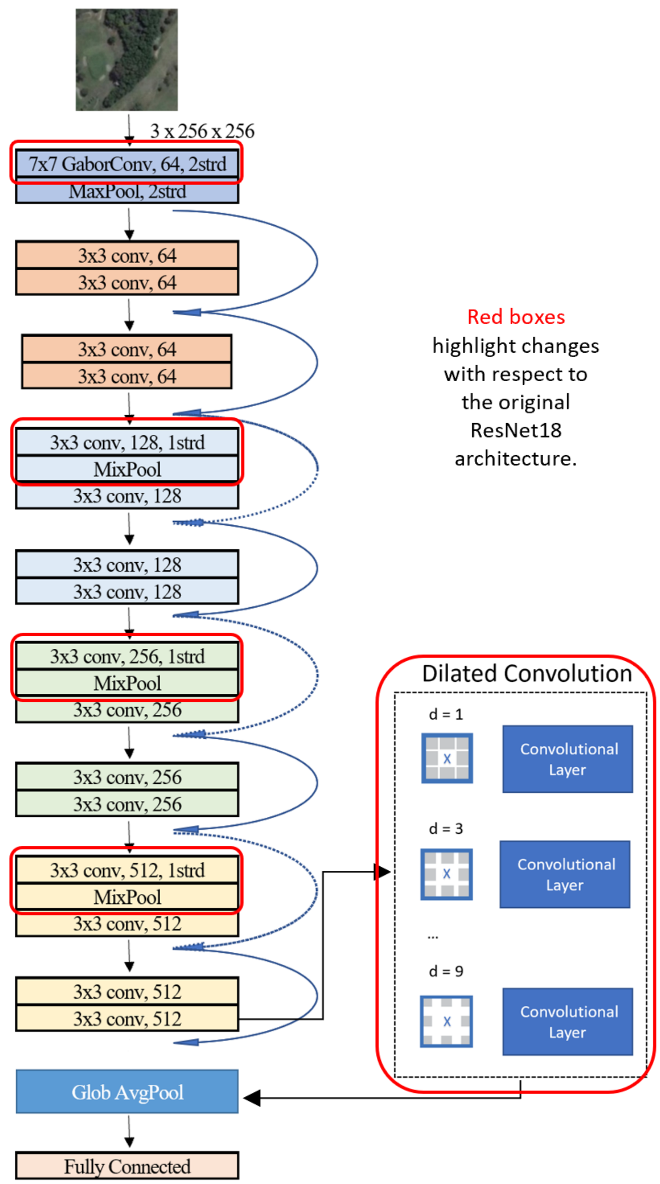

Figure 3. Proposed model: implemented a Gabor convolutional layer, residual blocks 2, 3, and 4: replaced stride 2 convolutional layers with 1 stride while added a mixture of maximum & average pooling layers, and added a dilated convolutional layer.

We applied the proposed solution described above on a ResNet18 architecture. In more detail, first we replaced the initial convolutional layer of the original ResNet18 architecture with a Gabor implementation (section 2.2.2). Then, we replaced stride 2 convolutional layers with stride 1 convolutional layer on residual blocks 2, 3, and 4, whilst adding after them, a mixture of maximum and average pooling layers (equation 4 with a scalar mixing portion $\alpha_l = 0.8$). Finally, we added a dilated convolutional layer prior to the global average pooling layer, resulting in the DilGabMPResNet18 architecture (figure 3).

2.3. Model Training

All models were trained for 100 epochs (based on when train/validation results stopped improving), using Adam Optimizer and Cross Entropy loss. For domain shift due to season, the dataset split was 80% for training and 20% for validation, while testing was performed on the opposite season data. For the mixed dataset, 80% of the data was split for train/validation purposes (out of which 80% train and 20% validation) and 20% for testing. Each experiment was run five times with different data splits. To tackle class imbalance and ensure classifiers were not biased towards the majority class, stratification was applied to the data. The only data augmentations applied during training, were geometric transformations, namely, rotation with 90% probability, horizontal and vertical flip with 50% and 10% probabilities respectively.

3. Results

3.1. State Of The Art (SOTA) Comparison

We first analyse the proposed dataset with a number of SOTA classifiers, specifically AlexNet [12], VGGNet [16], and ResNet [9] as a baseline for future research.

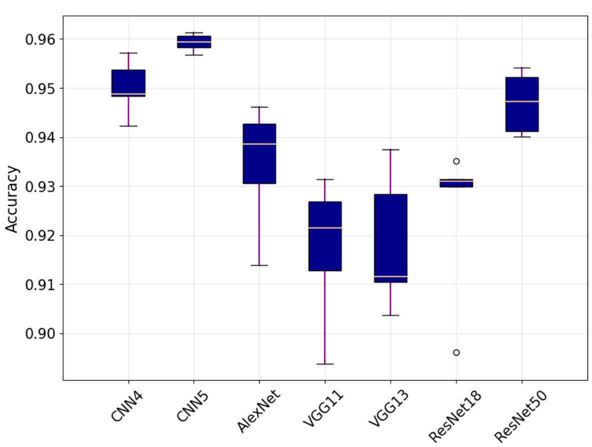

Figure 4. SOTA Model Comparison on Mixed Season Data.

The results are shown in figure 4. When using the proposed dataset to fully train deep convolutional neural network architectures such as ResNet, we see a performance drop, compared to shallower architectures. This is because deeper architectures, comprised of multiple layers, contain millions of parameters to be learned from the training data and thus require large datasets. Since the current dataset is comprised of tens of thousands of images, shallower networks like the five convolutional layer CNN5 have better performance.

A popular solution to tackle the above problem is to employ transfer learning methods where the model is pretrained in a large dataset like ImageNet, the weights of the lower layers are frozen, and the weights of the higher layers are optimised using the proposed dataset. However, when using this method with VGG or AlexNet, we observed a performance drop (between 0.1 to 1%) which perhaps highlights the difference between object and scene databases. The vast majority of ImageNet images are comprised of objects, often in an indoor setting. In contrast, the proposed dataset is comprised of outdoor scenes, taken from above. This emphasises the need for validated datasets of outdoor images. Given the performance of CNN5, we therefore use it as the comparator model in the subsequent results.

3.2. Feature Analysis

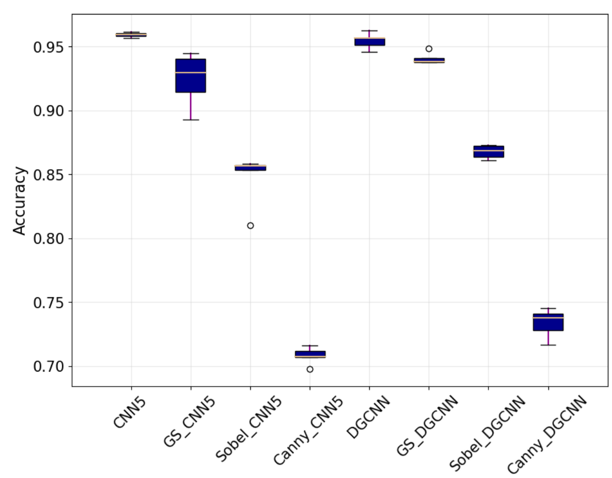

Figure 5. Feature Bias Performance on Mixed Season Data. For each architecture, the first part describes the data (i.e. grayscale (GS), edges with some texture (Sobel), and edges (Canny). The second part describes the architecture under review. E.g. GS_CNN5 is a CNN5 architecture trained/tested on grey-scale images.

We next compare the performance of CNN5 with our DGCNN model. Because we are interested in how robust performance is to seasonal domain shifts, we tested the models after altering colour and texture information (which varies greatly between seasons) to assess dependence on these features.

Figure 5 shows that removing colour features negatively affects both models but not catastrophically. While CNN5 has a slightly better performance overall compared to DGCNN, performance drops by approximately 2% when we remove colour, while DGCNN has a smaller drop in performance. When texture information is also absent (Canny edges), the performance drops significantly in both architectures, though the DGCNN architecture is again less affected, suggesting it extracts more information from shape features. By adding some texture features back in (Sobel), performance improves in both architectures. It thus seems that the CNN5 model relies slightly more on colour and texture features than the DGCNN architecture. This is important because when there is a seasonal domain shift, the change in flora mostly affects the colour (green) but also the texture features of the images.

3.3. Seasonal Data

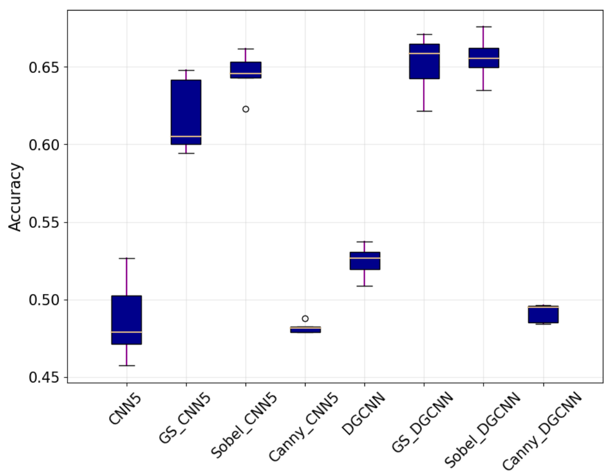

Figure 6. Train: Wet Season Data, Test: Dry Season Data Feature Bias Performance. The proposed architecture outperforms the original by 6% on colour and by 10% on gray scale images.

As a fuller test of robustness to seasonal domain shift, the classifiers were first trained using wet season data and tested with dry season data. This is to replicate issues that might occur if for instance a dataset is recorded at a single point in time, or the same point each year, but then needs to predict data from another point in time. As with the mixed data set, we also trained classifiers with grey-scale, Sobel and Canny-edge filtered images to see the reliance on colour/texture/shape.

As shown in figure 6, classifier performances for full images dropped compared to results for mixed season data. CNN5 accuracy drops from 96% to 49%, while DGCNN is a little more resilient to the domain shift with an accuracy drop from 95.6% to 54%. This resilience might have been predicted from figure 5 where DGCNN was slightly less detrimentally affected by the removal of colour than CNN5. However, while on that dataset there was only a small decrease in performance when removing colour, the big drop seen here indicates the issues with having a dataset biased to one season only.

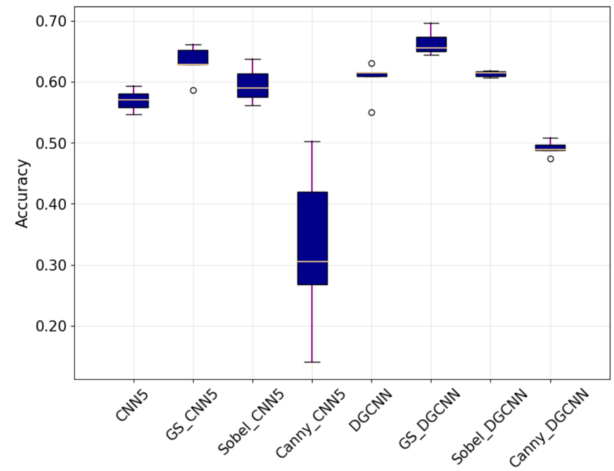

Figure 7. Train: Dry Season Data, Test: Wet Season Data Feature Bias Performance. The proposed architecture outperforms the original by 6% on colour and by 10% on grey-scale images.

The over-reliance on colour information in both models can be seen by training and testing using grey-scale images when the performance increases for both architectures (CNN5: 61% and DGCNN: 66% respectively). This is expected since the wet season causes flora to flourish, resulting in green colours dominating earth scenes, while in the dry season the lack of flora results in a brown colour. However, both models are still much below the performance on the mixed dataset, emphasising that care needs to be taken to avoid having data from one season only when training a model.

It is evident that colour is not the only feature that changes between seasons and that texture is also important when we use Sobel-filtered data performance increases further because flora growth affects texture as well as colour. Nonetheless, it is clear that some texture is important for specifying classes as removing all texture by applying a Canny edge filter, leads to a big performance drop for both models. However, compared to CNN5, DGCNN extracts better features from shape, having 3% better accuracy. Indeed across each of the different training sets, the DGCNN architecture outperforms CNN5, reinforcing the performance on the mixed dataset and giving greater robustness to seasonal change.

For a complete picture, we next train classifiers on the dry season data and test them on wet season data. Results are similar to when classifiers were trained using wet season data and predicted on the dry season but with some interesting differences. As shown in figure 7, compared to the mixed season data, performance drops significantly, but not quite as much as when trained on wet season data (figure

6). Additionally, as before, the DGCNN architecture outperforms CNN5 for each matched configuration, indicating its increased robustness to domain shift. However, the most notable change is that now classifiers trained with colour are not as hampered by the domain shift and indeed perform similarly to the Sobel filtered images, with the best performance when removing colour only. This is likely due to the more uniform colour range of dry-season images in which - especially for natural scenes - features on the ground are highlighted as textural differences (e.g. Barren Land in figure 1).

3.4. DilGabMPResNet18 Architecture

To assess the generality of the proposed architecture, we implemented it in a deeper, more complex model and train/tested it on larger more diverse datasets. We first assessed how each of the changes influenced the model's performance on mixed season data, and then how the combination of the proposed solutions affects the model's performance on domain-shifted data.

3.4.1 Ablation Study

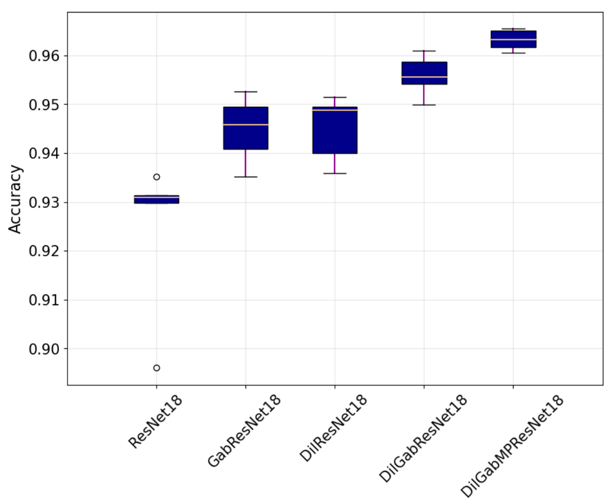

Figure 8. Models under review (from left to right): Original ResNet18, Gabor convolutional layer ResNet18, Dilated convolutional layer ResNet18, a combination of Gabor and dilated convolutional layer ResNet18, and a combination of Gabor, a mixture of max/average pooling layers and a dilated convolutional layer ResNet18 architectures trained on colour Mixed Season Data. The proposed architecture boosts performance by 3.7%.

Here we analyse the impact of each individual change to the ResNet18 architecture on the classification of the Mixed-season dataset, as well as combined changes. For the first test, we replaced the initial convolutional layer of ResNet18 with a Gabor convolutional layer (in which Gabor parameters could change during backpropagation). As seen in figure 8, adding a Gabor convolutional layer boosted the model's performance by 1.8% compared to ResNet18. Likewise, adding a dilated convolutional layer between the last convolutional layer and the global average pooling layer improved performance by 1.8% compared to standard ResNet18. However, the combination of an initial Gabor layer with a dilated convolutional layer together provided a larger performance increase of 2.7% than each individually (figure 8). The final change is to replace the 2-stride convolutional layers in residual blocks 2, 3, and 4 with 1 stride convolutional layer and the addition of a mixture of maximum & average pooling layers as per equation 4, which increased model performance by a further 1%.

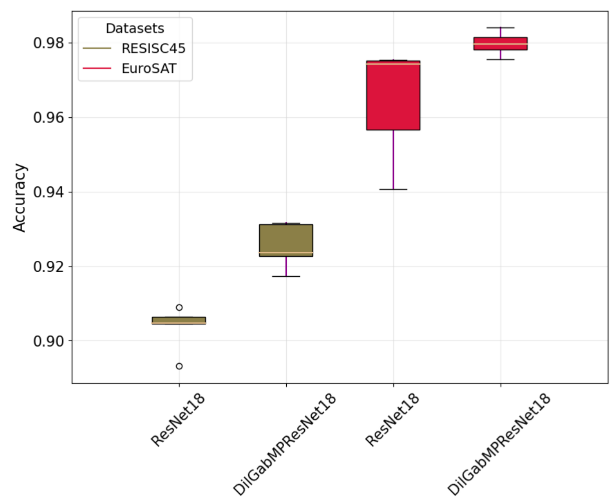

Figure 9. Original Resnet18 and Gabor convolutional with a mixture of max/average pooling layers and a dilated convolutional layer trained/tested on (gold) RESISC45, (red) EuroSAT, remote sensing datasets. The proposed architecture improves performance by 2% on the RESISC45 and by 1% on the EuroSAT datasets.

To validate the classifier's performance on other benchmark scene classification datasets, we assessed performance on the Remote Sensing Image Scene Classification (RESISC45) [4] and EuroSAT [10] datasets. RESISC45 is a large dataset comprised of 31,500 images covering 45 scene classes with 700 images per class. Moreover, the images vary in translation, spatial resolution, viewpoint, object pose, illumination, background, and occlusion. EuroSAT is based on Sentinel-2 satellite images, covering 13 spectral bands and is comprised of 27,000 labelled and geo-referenced images, covering 10 scene classes. On the RESISC45 dataset, the proposed model improves performance by 2% compared to the original ResNet18 architecture while on the EuroSAT dataset, the improvement is 1% (figure 9).

Table 2. Prediction accuracies (%) for domain adaptation experiments on the Office-Home dataset. Art (A), Clipart (C), Product (P), Real-World (R). A→C implies A is the source and C is the target. The proposed model outperforms the original one in all domain shift variations.

Architecture	R→P	R→C	R→A	P→A	P→C	P→R	A→C	A→P	A→R	C→A	C→P	C→R
ResNet18	21.92	9.32	13.14	5.89	6.71	15.74	4.77	5.59	11.02	4.99	6.42	6.61
DGMPResNet18	34.53	12.30	18.09	8.74	8.77	20.95	5.80	8.40	16.00	9.68	23.18	12.28

3.4.2 Domain Shifts

When trained on Wet and tested on Dry season data, or vice versa, the proposed architecture improves classification performance by 13% compared to the original ResNet18 (figure 10) showing that the new model works well in the presence of domain shifts. As a further test of the proposed model's performance on domain adaptation datasets, we applied it to the Office Home dataset proposed by Venkateswara et al. [18]. This is a domain adaptation dataset, comprised of 15,500 images, that covers four domains (Art, Clipart, Product and Real-World), where each domain consists of 65 categories. We tested both the original ResNet18 and the proposed architecture all combinations of source and target domains with results showing in table 2. The proposed solution outperforms the original architecture in all scenarios.

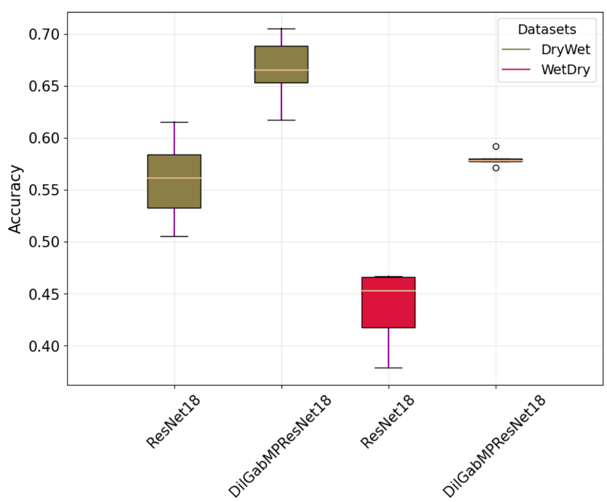

Figure 10. Original Resnet18 and Gabor convolutional with a mixture of maximum/average pooling layers and a dilated convolutional layer trained with (gold) Dry and tested on Wet, (red) Wet and tested on Dry, colour Season Data. The proposed architecture boosts performance by 13% in seasonal domain-shifted data.

Since the current work investigates domain shifts in seasonal variations, the results on similar domains like between the images of real-world (R) source, and product (P) target domain, or clipart (C) source, and product (P) target domains, where the proposed model increased performance by 14% and 17% is of interest.

Taken together with the improved performance on our seasonal domain shift dataset, these results indicate that the combination of extracting salient domain invariant features with a greater receptive field is beneficial for domain shifts.

4. Conclusion

Flora transformations between wet and dry seasons cause significant variation in aerial images in terms of shifts in colour and texture information which can hamper efforts to classify land use. As current datasets lack the ability to allow experimentation with domain shifts during seasonal variations, we here present a dataset that consists of wet and dry season data. We demonstrate the performance of various standard architectures on our dataset and, to highlight how features can change between seasons, we investigated how feature biases impact classifier performance and showed that biasing to texture and away from colour could improve performance in the presence of domain shifts. This led us to propose an architecture which uses a Gabor convolutional layer to extract salient features together with a larger receptive field. We show that classifiers trained on aerial images tend to rely more on texture, than other features (colour, shape) and that our architecture is a bit less reliant on colour features suggesting it will be more resilient to seasonal domain shifts. Finally, we applied similar changes to a deeper architecture and showed that this improved classification performance both on our mixed dataset and two other aerial image data sets, as well as improving performance in the presence of domain shifts.

References

[1] João Carrilho and Jorge Trindade. Sustainability in peri-urban informal settlements: A review. *Sustainability*, 14, 2022. 1

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014. 4

[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 4

[4] Gong Cheng, Junwei Han, and Xiaoqiang Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10):1865–1883, 2017. 7

[5] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6172–6180, 2018. 2

[6] Lijun Ding and Ardeshir Goshtasby. On the canny edge detector. *Pattern Recognition*, 34(3):721–725, 2001. 3

[7] Kunihiko Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982. 2

[8] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2018. 2

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4, 5

[10] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019. 7

[11] Katherine L Hermann, Ting Chen, and Simon Kornblith. The origins and prevalence of texture bias in convolutional neural networks. *arXiv preprint arXiv:1911.09071*, 2019. 2

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. 1, 5

[13] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013. 4

[14] Fiona Marshall, Linda Waldman, Hayley MacGregor, Lyla Mehta, and Pritpal Randhawa. On the edge of sustainability: Perspectives on peri-urban dynamics. *STEPS Working Paper*, 35, 2009. 1

[15] Ivan Adriyanov Nikolov, Mark Philip Philipsen, Jinsong Liu, Jacob Velling Dueholm, Anders Skaarup Johansen, Kamal Nasrollahi, and Thomas B Moeslund. Seasons in drift: A long-term thermal imaging dataset for studying concept drift. In *Thirty-fifth Conference on Neural Information Processing Systems*, 2021. 2

[16] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5

[17] Irwin Sobel and Gary Feldman. A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in*, pages 271–272, 1968. 3

[18] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5018–5027, 2017. 8

[19] Georgios Voulgaris, Andrew Philippides, and Novi Quadrianto. Deep learning robustness to domain shifts during seasonal variations. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 417–420. IEEE, 2022. 4

[20] Junjue Wang, Zhuo Zheng, Ailong Ma, Xiaoyan Lu, and Yanfei Zhong. Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation. *arXiv preprint arXiv:2110.08733*, 2021. 2

[21] Shenlong Wang, Min Bai, Gellert Mattyus, Hang Chu, Wenjie Luo, Bin Yang, Justin Liang, Joel Cheverie, Sanja Fidler, and Raquel Urtasun. Torontocity: Seeing the world with a million eyes. *arXiv preprint arXiv:1612.00423*, 2016. 2

[22] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S Huang. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7268–7277, 2018. 4

[23] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. Aid: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3965–3981, 2017. 2

[24] Gui-Song Xia, Wen Yang, Julie Delon, Yann Gousseau, Hong Sun, and Henri Maître. Structural high-resolution satellite image indexing. In *ISPRS TC VII Symposium-100 Years ISPRS*, volume 38, pages 298–303, 2010. 2

[25] Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 270–279, 2010. 2

[26] Qin Zou, Lihao Ni, Tong Zhang, and Qian Wang. Deep learning based feature selection for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, 12(11):2321–2325, 2015. 2