Cross-Domain Generalization of CNN-based Image Classification Problems using Wavelet Transforms and Applications in MRI

Mid-Semester Evaluation
(Wavelets, EE678 - Autumn 2024)

Submitted By
Group 22
Abhijat Bharadwaj (210020002)
Animesh Kumar (21D070012)
Uvesh Mohammad (21D070084)

Submitted To
Prof. Vikarm M. Gadre
Dept. of Electrical Engineering
Indian Institute of Technology Bombay
Mumbai, Maharashtra, India

Contents

1 Question (a): Basic Information

The report is submitted by *Group 22* consisting of the following students:

- Abhijat Bharadwaj (210020002)
- Animesh Kumar (21D070012)
- Uvesh Mohammad (21D070084)

Title of the work: "*Cross-Domain Generalization of CNN-based Image Classification Problems using Wavelet Transforms and Applications in MRI*"
For references, see section "References".

2 Premise: Convolutional Neural Network

A Convolutional Neural Network (CNN) is a deep learning model designed to process data with a grid-like structure, such as images. CNNs are widely used in image recognition, video analysis, and other tasks that involve spatial data, where they excel at capturing patterns like edges, textures, and shapes through their specialized layers.

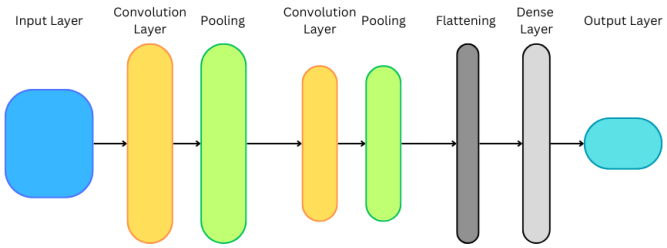

Fig. 1: Model architecture of a convolutional neural network

Fundamentally, CNNs are based on the mathematical operation of convolution, which is often denoted by

$$(f*g)(x) = \int_{-\infty}^{+\infty} f(t)g(x-t)\mathrm{dt} \quad (1\mathrm{D}) \tag{1}$$

$$(f*g)(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(u,v)g(x-u,y-v)\mathrm{dudv} \quad (2\mathrm{D}) \tag{2}$$

where $f(.)$ and $g(.)$ are functions of an independent variable x and t is an intermediate variable. We can notice that the operator is inherently non-linear. The same can be extended to multi-dimensional functions. CNNs apply filters (or kernels) to the input data, which are small, learnable matrices that slide over the input data, performing a convolution operation to detect local patterns such as edges, textures, or shapes. The convolution layers are followed by pooling and fully connected layers (also called dense layers) to complete the CNN.

2.1 Need for improvement

Classical CNNs perform tremendously well in learning features from one dataset type, but to make the design scalable and generalizable, CNN architectures involve numerous hyperparameters that must be fine-tuned. Although transfer learning is helpful, there is still a need for more precise design principles [1].

3 Question (b): Wavelets in CNN

Question (b): Explain the connection(s) between wavelets/ filter banks and convolutional neural networks/ machine learning structures/ deep learning structures that you have explored. Bring out specifically, how wavelets have come into the realm of convolutional neural networks/ machine learning structures/ deep learning structures in the reference(s) that you have studied, in some depth.

Cross-domain generalization is the ability of a learning algorithm to learn from one or more domains and apply that knowledge to other domains. A practical interpretation of the problem can be explained through the following example — consider a set of brain tumor images obtained from scanner S_1 using protocol T_1. These images might possess specific properties influenced by the medium of acquisition. A classifier CNN trained on this given dataset might learn these influenced features. Hence, the classification ability of that CNN will suffer if it is used on an MRI image from a different scanner or protocol. The authors of [1] utilize the **Wavelet Scattering Convolution Network**, developed by *J. Bruna* and *S. Mallat* [2], and isometry equivariant non-expansive operators (IENEOs), from work in [3], in an attempt to develop an analytical version of the CNN, improving the cross-domain generalization ability.

3.1 Mathematical Modelling

Cross-domain generalization aims to develop models that are robust to the distribution shift. Identifying features *invariant* and *equivariant* to shifts becomes helpful to achieve the same. CNNs perform their functions by learning a mathematical function based on the training data. Because of CNN's architecture, these functions $f(x)$ are treated as black boxes. We aim to break the CNN using an interpretable intermediate internal representation $\phi(x)$ for all data points x using wavelet scattering coefficients and IENEOs in the A-CNN architecture. Let us define some symbols below before we move forward:

X **:** all data points
$x \in X$ **:** a data instance in X
D **:** set of all class labels (outputs of classifier)
$f : X \to D$ **:** conventional deep CNN (learned) classifier (black box)
Φ **:** set of all internal representations in A-CNN

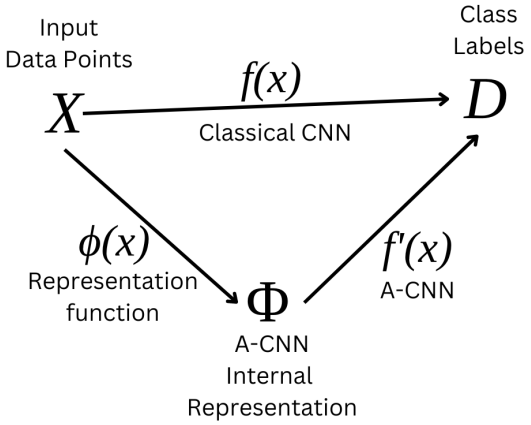

Fig. 2: [1, Fig. 1] Function Map of A-CNN setup

$\phi(x)$ **:** internal representation of x in A-CNN

$\phi : X \to \Phi$

$\Phi = \{\phi(x),\ \forall x \in X\}$

f' **:** analytical classifier learned by A-CNN

$f' : \Phi \to D$

G **:** group of global symmetries acting on X

$g \in G$ **:** group of symmetries operating on function f

$g(x)$ **:** transformed image of x by applying group transformation

3.2 Invariance

$\phi_i(x)$ is said to be an *invariant* representation [1] of an object $x \in X$ with respect to some group G if and only if

$$\phi_i(x) = \phi_i(g(x)) \quad \forall g \in G \tag{3}$$

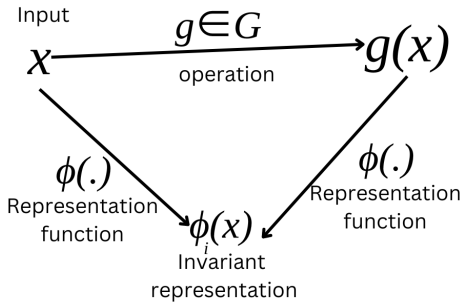

Fig. 3: Invariant Representation

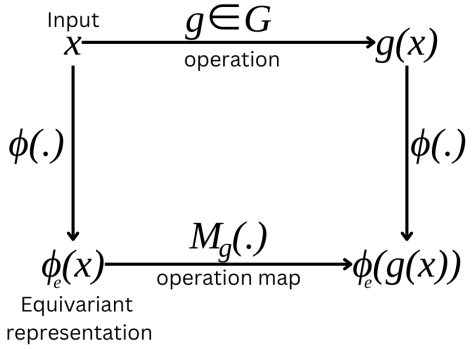

Fig. 4: Equivariant Representation

3.2.1 Invariance and Cross-Domain Generalization

Features invariant to some group G would appear the same in different domains if the domains differ by some combination of operations in G. This would help us learn common features from images between the domains, i.e., we don't need techniques like cross-validation to ensure noise isn't being learned.

3.3 Equivariance

$\phi_e(x)$ is said to be an *equivariant* representation [1] of an object $x \in X$ with respect to some group G if and only if

$$\exists_{M_g} : \phi_e(g(x)) = M_g(\phi_e(x)) \quad \forall g \in G \tag{4}$$

Here, M_g is a map that translates the effects of operations g operating on X to $\phi(x)$.
Please note that some literature refers to the about property as *covariance* and equivariance is treated as the particular case of covariance such that $M_g(.) = g(.)$. But since the main work under analysis [1] uses the prior definition, we will stick to the same throughout the report.

3.3.1 Equivariance and Cross-Domain Generalization

Features equivariant to some group G would appear symmetrically transformed in different domains if the domains differ by some combination of operations in G. These features reflect how the objects being learned are transformed across various domains. The existence of an $M_g \forall g \in G$ implies that the equivariant features can be translated to a standard representation without loss of information.

3.4 Wavelet Scattering Transform: Achieving Invariance

The standard method of obtaining a *representation* is first to obtain the *basis* of the *vector space* in which we want the representation to be obtained. Standard signal processing techniques use the Fourier Series, which uses Fourier sinusoidal waves as basis vectors. But unlike Fourier sinusoidal waves, as Wavelets are localized waveforms, they are much more stable to deformations [2]. A wavelet transform computes convolutions with translated and rotated wavelets.

$$f(u,v) = \psi(u,v) * x(u,v)$$
$$\implies \psi(u-u_1, v-v_1) * x(u-u_2, v-v_2) = f(u-u_1-u_2, v-v_1-v_2)$$

Since convolution commutes with translation, it is equivariant to translation and not invariant. Thus, we shift to scattering transform.

3.4.1 Two-dimensional Directional Wavelets

Let $G_{rotations}$ be a group of rotations r of angles $2k\pi/K$ for $0 \leq k \leq K$. Two-dimensional directional wavelets are obtained by rotating a single band-pass filter ψ by $r \in G_{rotations}$ and dilating it by 2^j for $j \in \mathbb{Z}$ [2]:

$$\psi_\lambda(u) = 2^{-2j}\psi(2^{-j}r^{-1}u) \quad \text{with } \lambda = 2^{-j}r \tag{5}$$

If the Fourier transform is centered at frequency η, then $\hat{\psi}_{2^{-j}r}(\omega) = \hat{\psi}(2^j r^(-1))\omega$ has support centred at $2^{-j}r\eta$ and a bandwidth proportional to 2^{-j}. The index $\lambda = 2^{-j}r$ gives the "frequency location" of ψ_λ [2]. Note, $|\lambda| = 2^{-j}$.
From here, the authors proceed with using "Morlet Wavelet" for further work as it resembles human perception [1] [4]. Morlet Wavelet ψ is given by:

$$\psi(u) = \alpha\left(e^{ju\zeta} - \beta\right)e^{-|u|^2/(2\sigma^2)} \tag{6}$$

where $\beta \ll 1$ is adjusted so that $\int \psi(u)\mathrm{du} = 0$. Its real and imaginary parts are nearly *Quadrature Phase Filters* [2].

3.4.2 Wavelet Transform

The wavelet transform of x is $\{x * \psi_\lambda(u)\}_\lambda$. Since it is a convolution, it is not translation invariant.
Introducing invariance: To introduce translation invariance, we will need non-linearity. For any linear or non-linear operator Q, $\int Qx(u)\mathrm{du}$ is always translation invariant. But $\int x * \psi_\lambda(u)\mathrm{du}$ gives trivial invariant ($= 0$). This motivates us to generate $\mathbb{L}_1$ *norm* as our translation coefficient [2]:

$$||x * \psi_\lambda||_1 = \int |x * \psi_\lambda(u)|\mathrm{du} \tag{7}$$

Validity of $\mathbf{L}_1$-norm as invariant: We have calculated the $\mathbf{L}_1$-norm of an image and another image where the object is spatially translated [Fig. 5]. The percentage change in the $\mathbf{L}_1$-norm of the image comes out to be around $\mathbf{0.0006}\%$. This minimal change suggests that the $\mathbf{L}_1$-norm effectively preserves translation invariance.
Recovering Information Loss: The $\mathbf{L}_1$ norms $\{||x * \psi_\lambda||_1\}_\lambda$ form a crude representation of sparsity of wavelet coefficients. The loss of information does not come from the modulus, which removes the complex plane of the wavelet transform. Rather, the integration of $|x * \psi_\lambda(u)|$ removes all non-zero frequency [2]. The non-zero frequencies are recovered by calculating the wavelet coefficients of $|x * \psi_{\lambda_1}(u)|$ ($= \{|x * \psi_{\lambda_1}(u)| * \psi_{\lambda_2}(u)\}_\lambda$). Thus, the $\mathbb{L}_1$ norm of the new coefficient using λ_1 and λ_2, $\||x * \psi_{\lambda_1}(u)| * \psi_{\lambda_2}(u)\|_1$, define a much larger family of invariants.

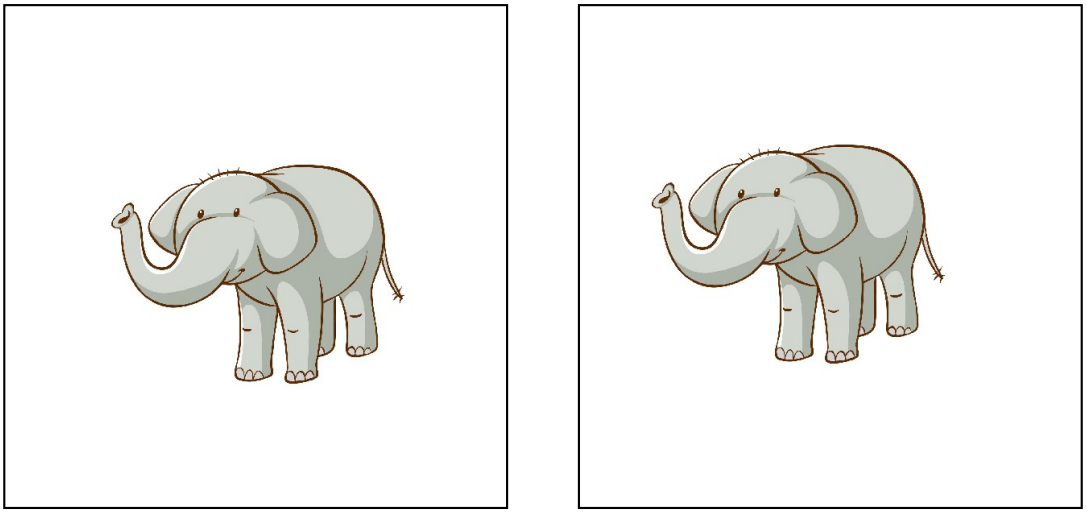

Fig. 5: (left) Original image; (right) Translational Shift of original image

3.4.3 Wavelet Scattering Transform

Define $U[\lambda]x = |x * \psi_\lambda|$. Along any *path* $p = (\lambda_1, \lambda_2, ..., \lambda_m)$, an ordered product of non-linear and non-commuting operators is calculated:

$$U[p]x = U[\lambda_m]...U[\lambda_2]U[\lambda_1] = |...||x * \psi_{\lambda_1}| * \psi_{\lambda_2}|... * \psi_{\lambda_m}| \tag{8}$$

with $U[\emptyset]x = x$. A *scattering transform* [2] along the path p is defined by:

$$\overline{S}x(p) = \mu_p^{-1} \int U[p]x(u)\mathrm{du} \quad \text{with } \mu_p = \int U[p]\delta(u)\mathrm{du} \tag{9}$$

Each Scattering coefficient $\overline{S}x(p)$ is invariant to a translation of x.

3.4.4 Windowed Scattering Transform

For classification, it's often useful to compute "localized descriptors," i.e., features stable against small translations up to a certain scale (2^J) while preserving differences at larger scales. This is done by applying a spatial window at scale 2^J denoted by:

$$\phi_{2^J}(u) = 2^{-2J}\phi(2^{-J}u), \tag{10}$$

focusing the scattering transform around the point of interest. It defines a *windowed scattering transform* in the neighbourhood of u:

$$S[p]x(u) = U[p]x * \phi_{2^J}(u) = \int U[p]x(v)\phi_{2^J}(u-v)\mathrm{dv}, \tag{11}$$

and hence,

$$S[p]x(u) = |...||x * \psi_{\lambda_1}| * \psi_{\lambda_2}|... * \psi_{\lambda_m}| * \phi_{2^J}(u), \tag{12}$$

where $S[\emptyset]x = x * \phi_{2^J}$. The averaging by ϕ_{2^J} implies that if $x_c(u) = x(u-c)$ with $|c| \ll 2^J$, then the windowed scattering is nearly translation invariant $S[p]x \approx S[p]x_c$.

Observing Effects of Wavelet Scattering Transform: We have applied zero-order, and first-order scattering transforms to a random input signal, as shown in Figure 6. Additionally, we processed an MNIST image labeled '0' [5] and two images from different classes of the Office-31 dataset [6]. The results are as follows:

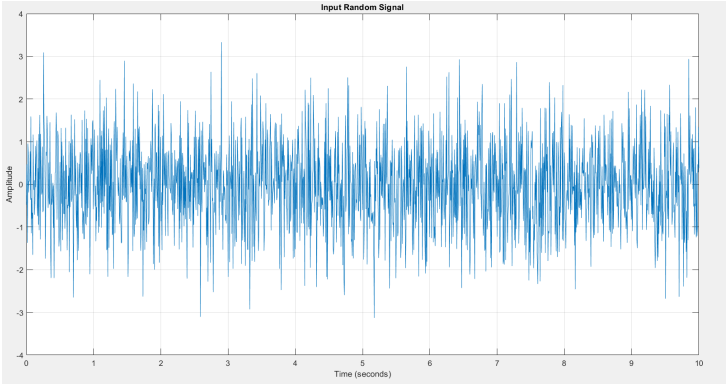

Fig. 6: Random Input Signal

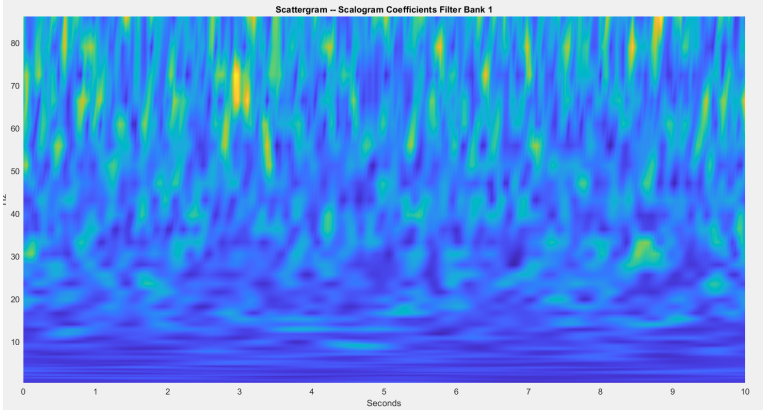

Fig. 7: Level 1 Scalogram Coefficients of Fig. 6

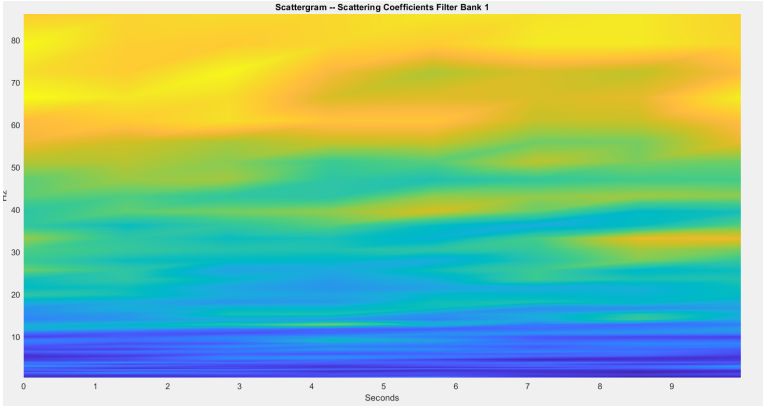

Fig. 8: Level 1 Scattergram Coefficients of Fig. 6

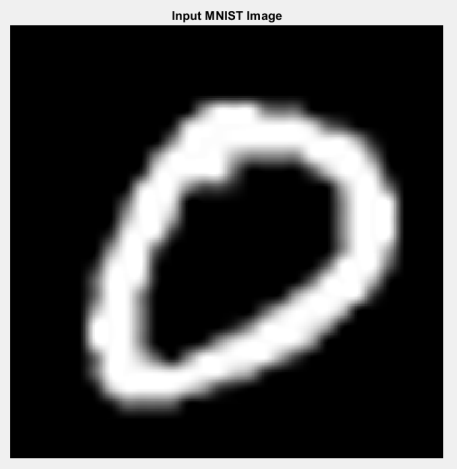

Fig. 9: Input MNIST image of label ‘0’

Fig. 10: Zero-order Scattering Coefficients

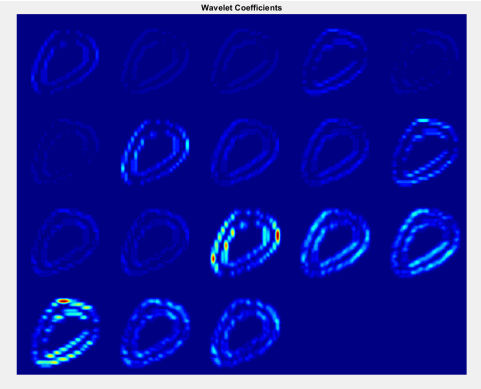

Fig. 11: Wavelet coefficients MNIST image of label ‘0’

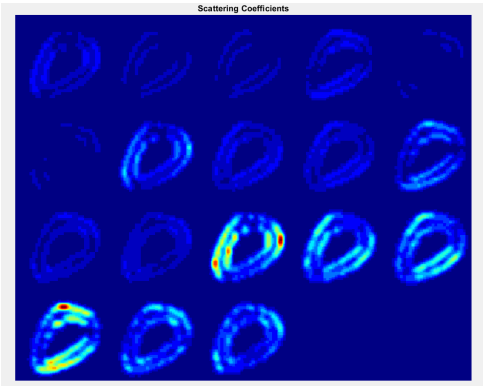

Fig. 12: Scattering coefficients MNIST image of label ‘0’

Fig. 13: Input image of office-31 from class monitor

Fig. 14: Zero-order Scattering Coefficients

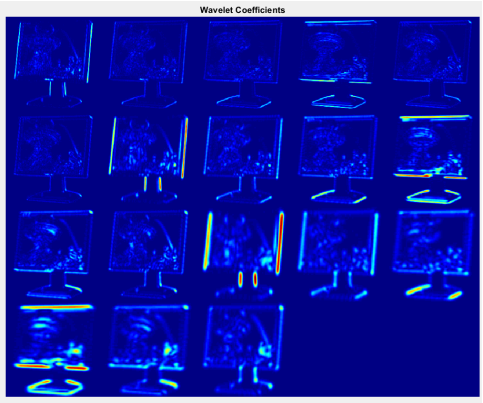

Fig. 15: Wavelet coefficients image of office-31 from class monitor

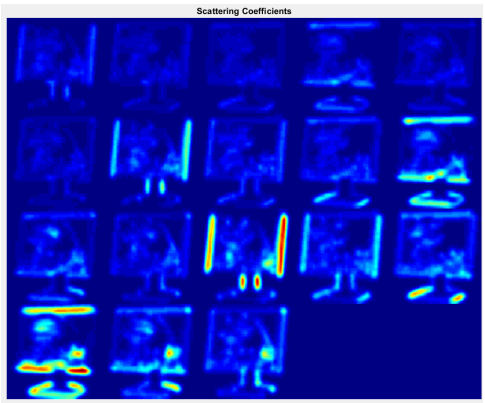

Fig. 16: Scattering coefficients image of office-31 from class monitor

Fig. 17: Input image of office-31 from class desktop

Fig. 18: Zero-order Scattering Coefficients

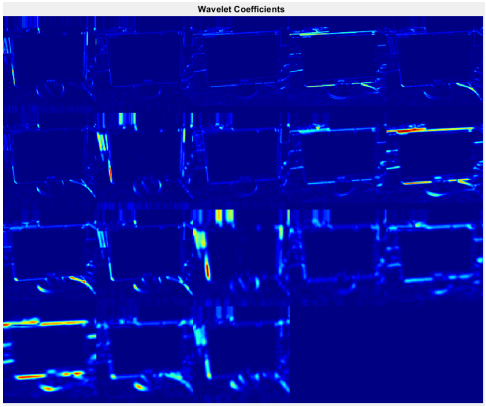

Fig. 19: Wavelet coefficients image of office-31 from class desktop

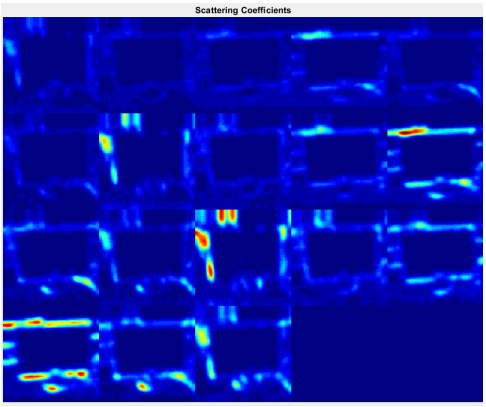

Fig. 20: Scattering coefficients image of office-31 from class desktop

3.5 IENEO: Achieving Equivariance

IENEOs are designed to preserve symmetries in image data. An IENEO is defined as:

$$H_p(i,j) = \sum_{m=1}^{k} a_m g_{\tau_m}(i^2 + j^2), \tag{13}$$

where g_{τ_m} represents Gaussian functions with varying widths τ_m, and a_m are coefficients. This operator is used to transform images while preserving their symmetry properties. The equivariant representation $\phi(i,j)$ is obtained by convolving the input image $x(i,j)$ with the IENEO [1]:

$$\phi(i,j) = \int_{\mathbb{R}} x(\alpha,\beta) \frac{H_p(i-\alpha, j-\beta)}{\|H_p\|_1} d\alpha\, d\beta \tag{14}$$

This convolution process ensures that if the input image undergoes a transformation, the resulting representation $\phi(i,j)$ maintains the same symmetry, thus preserving the equivariance. The equivariant nature of convolution is already proven in Section 3.4. The choice of IENEOs is guided by persistent homology [3], which evaluates and selects operators that best enhance class separation and reduce intra-class variation.

3.6 The A-CNN Architecture

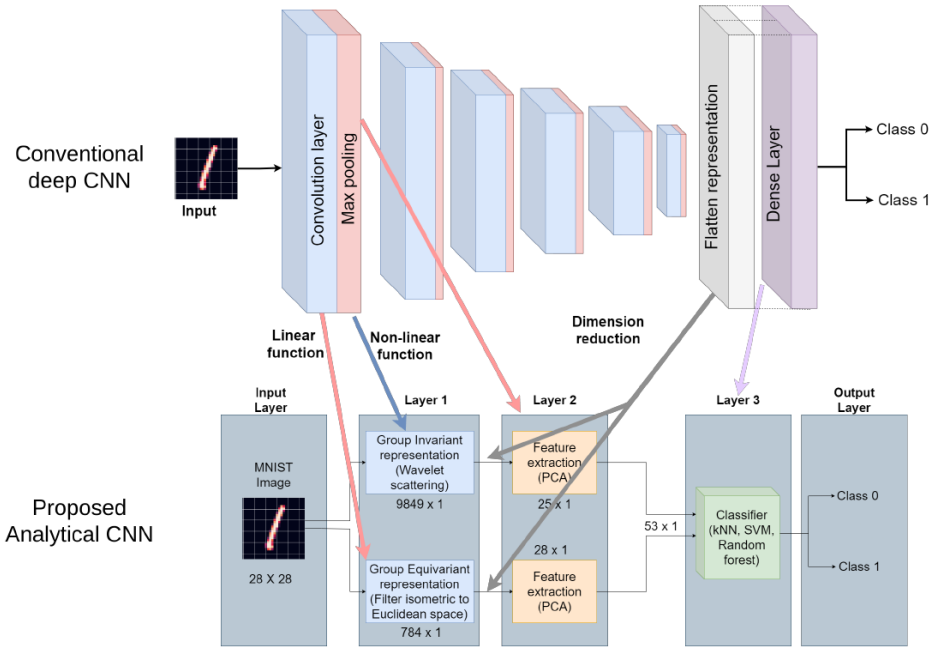

Fig. 21: [1, Fig. 2] The proposed analytical CNN architecture employing wavelet scattering

The A-CNN architecture in the original work [1] proposes to sequentially replace the layers of conventional CNN using a functionally equivalent layer. The new layers and their correlation to the conventional CNN layers are displayed in [Fig. 21].

3.6.1 Layer 1

In the proposed A-CNN, the first layer focuses on generating robust image representations by employing Scattering Wavelet Transform (SWT) and group equivariant operators. SWT is used to achieve translation invariance, ensuring that the representation of an image remains unchanged despite translations. This replaces the conventional convolutional layer of CNNs, which learns filters through training to detect features. Instead of learning these filters, the A-CNN uses predefined wavelets and operators to directly extract invariant and equivariant features.

3.6.2 Layer 2

The second layer performs dimension reduction using Principal Component Analysis (PCA). PCA compresses the feature set obtained from the first layer, retaining the most significant aspects while reducing dimensionality. This layer substitutes the pooling and flattening operations of traditional CNNs, which reduce the size of feature maps and prepare them for classification. PCA offers a parameter-free approach to feature reduction, simplifying the data while preserving key information.

3.6.3 Layer 3

The third layer involves classification using analytical models such as k-Nearest Neighbors (kNN), Support Vector Machine (SVM), or Random Forest (RF). These classifiers operate on the compact feature set produced by PCA. This replaces the fully connected layers found in traditional CNNs, which are used for classification after feature extraction.

4 Questions (c,d): Economy and Interpretability

4.1 Question (c): Definition

What does economy mean in the context of machine learning/ deep learning/ neural networks? What does explainability/ interpretability mean in the context of machine learning/ deep learning/ neural networks?

4.1.1 Economy

In the context of machine learning, deep learning and neural networks, *economy* refers to the efficient use of computational and resource-related aspects of model development and deployment. It emphasizes balancing performance with the constraints of time, memory, energy, and storage. The economy encompasses several factors that contribute to the efficient use of resources while maintaining model performance. Some of them are listed below:

- Computational Efficiency
- Training Time

- Memory Usage
- Energy ConsumptionTime
- Inference Speed
- Scalability

4.1.2 Explainability and Interpretability

Explainability or *interpretability* refers to the degree to which the internal mechanics of a machine learning model can be understood. It involves making the model's behavior transparent and understandable to users, especially non-experts. It aims to answer the question, "Why did the model make a particular decision?" Since deep neural networks consists of many layers and millions of parameters, it's challenging to decipher how specific inputs are transformed into outputs. It's one of the examples of a black-box models. In black box models, explainability involves techniques that provide insights into the model's decision-making process after the fact. Since the model's internal workings are not transparent, explainability methods attempt to uncover or approximate how the model arrived at its predictions.

There is often a trade-off between model complexity and interpretability. More complex models (like deep neural networks) can capture intricate patterns in data but are more challenging to interpret. In contrast, simpler models (like linear regression) are easier to understand but might not capture complex relationships as effectively.

4.2 Question (d): Benefits of the A-CNN Architecture

Explain the benefits that have accrued, from this union of wavelets/ filter banks and convolutional neural networks/ machine learning structures/ deep learning structures. In particular, emphasize and identify the benefits in terms of economy and/or explainability/ interpretability.

4.2.1 Benefits to CNN Economy:

- **Wavelet-Based Feature Extraction Removing Weight Learning:** The A-CNN leverages predefined wavelets (such as Morlet wavelets) to extract features. Since a predefined class of wavelets completes the basis for representation vector space, the features (invariant) can be learned without requiring filter learning through backpropagation, effectively removing the whole weight-training process. This reduces the computational cost compared to conventional CNNs, making the model more efficient.
- **Compact Representation with Minimal Information Loss:** The model employs Principal Component Analysis (PCA) to reduce the dimensionality of the extracted wavelet-based invariant and equivariant representations. This reduces the computational overhead while retaining meaningful features.
- **Simplified Architecture:** Unlike traditional CNNs with deep hierarchies, the A-CNN uses only three main layers, reducing the model's overall complexity and training time.

This also indicates the ability of the A-CNN architecture to extract the features without using multiple layers.

- **Scope for Parallel Computation:** The A-CNN architecture also enables a scope for parallel extraction of features. So, in a multi-threaded or multi-processor architecture, parallel processing can be enabled cutting the computation time in half.

4.2.2 Benefits to Explainability and Interpretability of CNN:

- **Structural Changes with Interpretable Intermediates**: In an A-CNN, interpretability is enhanced throughout the network due to its structured design, which uses mathematically defined transformations like wavelets and equivariant/invariant features. Thus, the output of each layer of the A-CNN provides vector(s) with physical significance.
- **Invariance and Equivariance from Layer 1**: Equivariant features in A-CNN preserve the effects of transformations (such as rotations), while invariant features remain constant under certain transformations (like translation). This allows you to directly observe how the model processes and reacts to input variations, making it clear how the network handles transformations like rotation or scaling.
- **Understandable Dimension Reduction from Layer 2**: After extracting these features, A-CNN applies PCA for dimensionality reduction, where the process is transparent, and you can see precisely which features are retained and why. Contrasting this with max-pooling, where the maximum value might not have any particular significance.
- **Reasoning for Classification from Layer 3**: In the final classification layer, A-CNN uses interpretable kNN, SVM, making decisions easy to trace. This contrasts with traditional CNNs, where dense layers and softmax obscure how complex features are learned and decisions are made. A-CNN allows interpretability to be injected at any point, enabling clear tracing of the decision-making process from input to output, unlike the black-box nature of traditional CNNs.

5 Question (e): Applications in Magnetic Image Resonance (MRI)

List and explain, how the ideas that you have explained in Questions (b) and (d) above play a useful role in the specific vertical/ theme that your group has chosen for its semester course project, if you have been able to identify the same

- **Robustness to Variations:** By utilizing wavelet transforms and equivariant techniques, A-CNNs effectively handle variability in MRI acquisition protocols and scanner types. This ensures that the detection of features like brain lesions or tumor boundaries remains accurate and reliable, regardless of differences in imaging equipment or orientation of the images.

- **Efficiency and Speed:** The use of wavelet-based features and a streamlined network architecture in A-CNNs reduces computational demands, leading to quicker processing of MRI images. This is essential for timely diagnostics, particularly in emergency situations or high-throughput screening scenarios where rapid analysis is critical. Also, since tomography is time-consuming, speeding up post-processing becomes significant.

- **Clinical Interpretability:** A-CNNs offer clear interpretability in MRI analysis by providing insights into how specific features, such as tumor characteristics or brain structure anomalies, influence diagnostic decisions. This transparency aids clinicians in understanding and trusting the automated results, facilitating smoother integration into diagnostic workflows and decision-making processes.

- **Feature Simplification for Better Segmentation:** By reducing high-dimensional feature space through autoencoding using Deep Wavelet Autoencoder (DWA) [7] on invariant wavelet features, it allows the model to distinguish between different regions (e.g., healthy vs abnormal tissue) more effectively by removing the influence brought by differences in training data. Moreover, as seen in Fig. 15, different order scattering coefficients represent various significant features, which will improve the segmentation decision making. We plan to explore this more in the *Evaluation Component-III*.

6 Question (f): Beyond the References

If you have been able to think beyond what the references have described and have come out with some of your own relevant ideas to bring economy/ explainability/ usefulness through the union of wavelets/ filter banks and machine learning/ deep learning/ neural networks, please explain the same, clearly and unambiguously.

Building on the above-explained A-CNN architecture from [1], we attempted to dive deeper and understand the properties of the invariant features obtained using Windowed Wavelet Scattering (Section 3.4.4). After experimenting with different datasets, we observed how the t-SNE plots of objects from the same class behave when data is collected from various domains. We then compared the same to the behavior of the calculated invariant features of the same objects and explored future scopes on how the invariant feature extraction can be exploited to help the neural networks beyond CNNs.

6.1 Experiments

To explore the union of wavelets and machine learning for enhancing explainability, cross-domain generalization, and practical utility, we conducted an experiment using Scattering Wavelet Transform (SWT) on images from different domains. Initially, we plotted the t-SNE of a single class across the three domains without SWT. The dataset used is Office-31, which

has three domains, namely, webcam, DSLR and amazon [6]. The results obtained are shown in the following plots. In [Fig. 22], we plotted t-SNE of the images in the different domains (webcam and DSLR) of the same label 'tape-dispenser' and t-SNE of the same images after the first level SWT. In [Fig. 23], the images used are of the class 'bottle' of domains webcam and amazon. Similarly, in [Fig. 24], [Fig. 25], [Fig. 26] and [Fig. 27], images used are of classes bottle, bike, desktop-computer and back-pack of all the three domains.

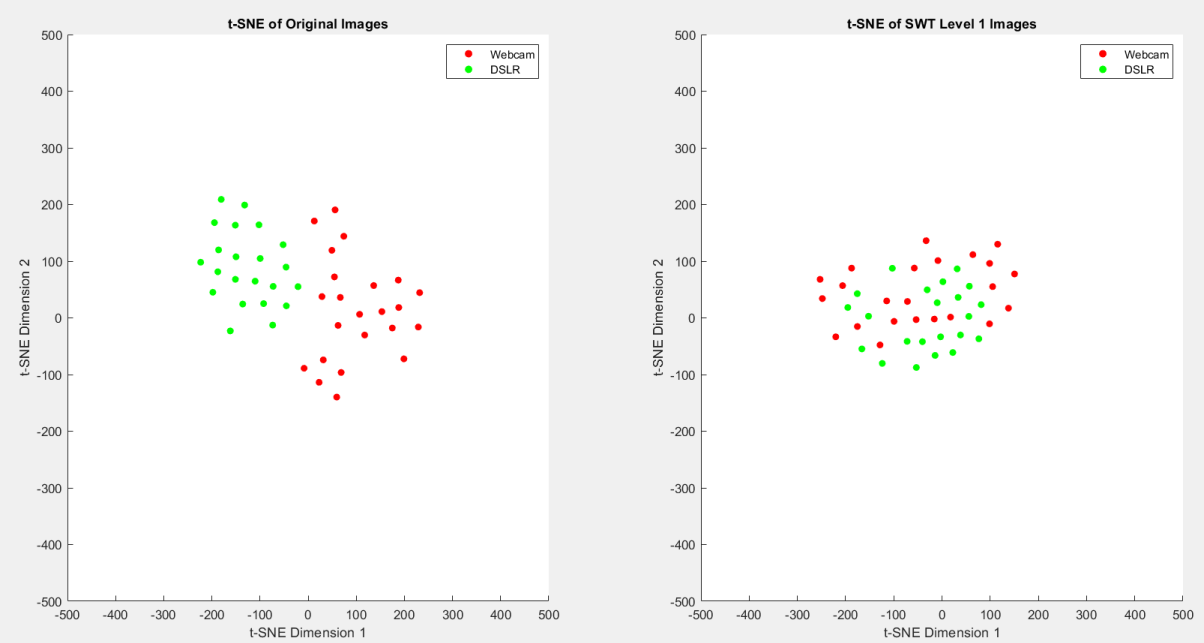

Fig. 22: t-SNE of images form class tape dispenser of domains webcam and dslr

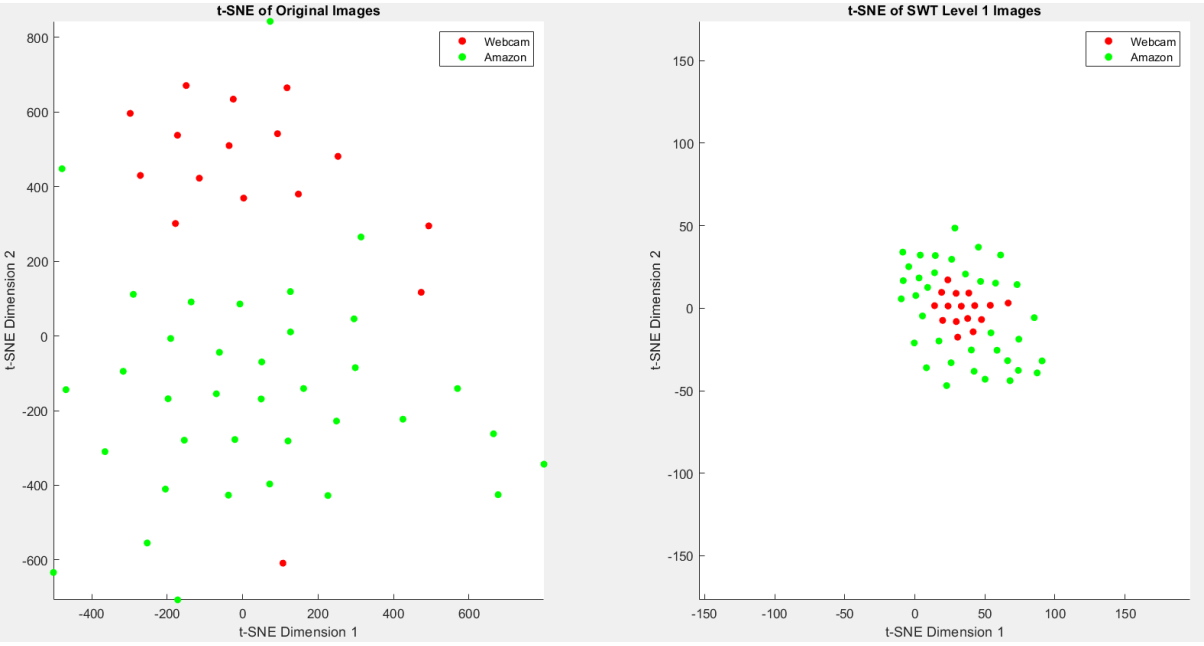

Fig. 23: t-SNE of images from class bottle of domains webcam and amazon

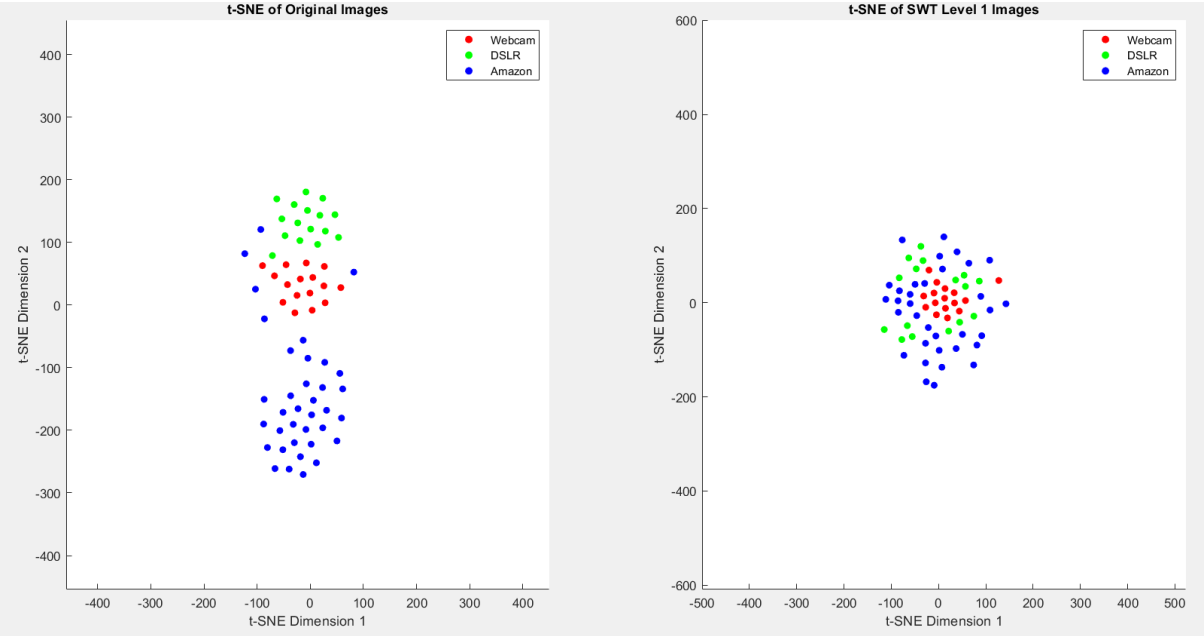

Fig. 24: t-SNE of images from class bottle of all three domains

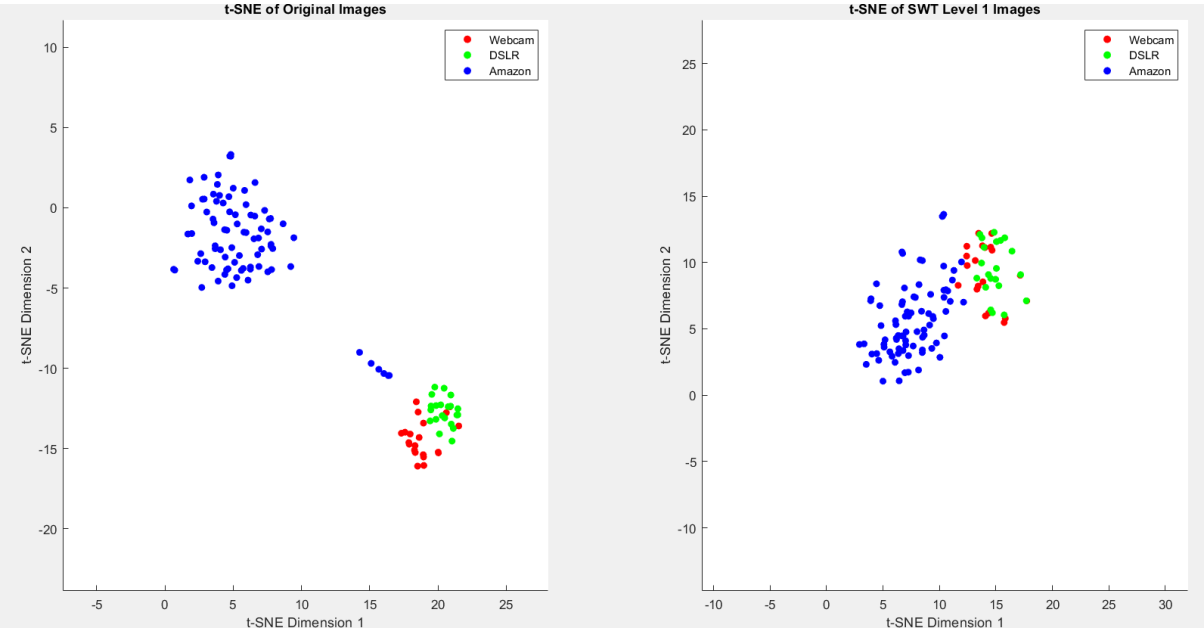

Fig. 25: t-SNE of images from class bike of all three domains

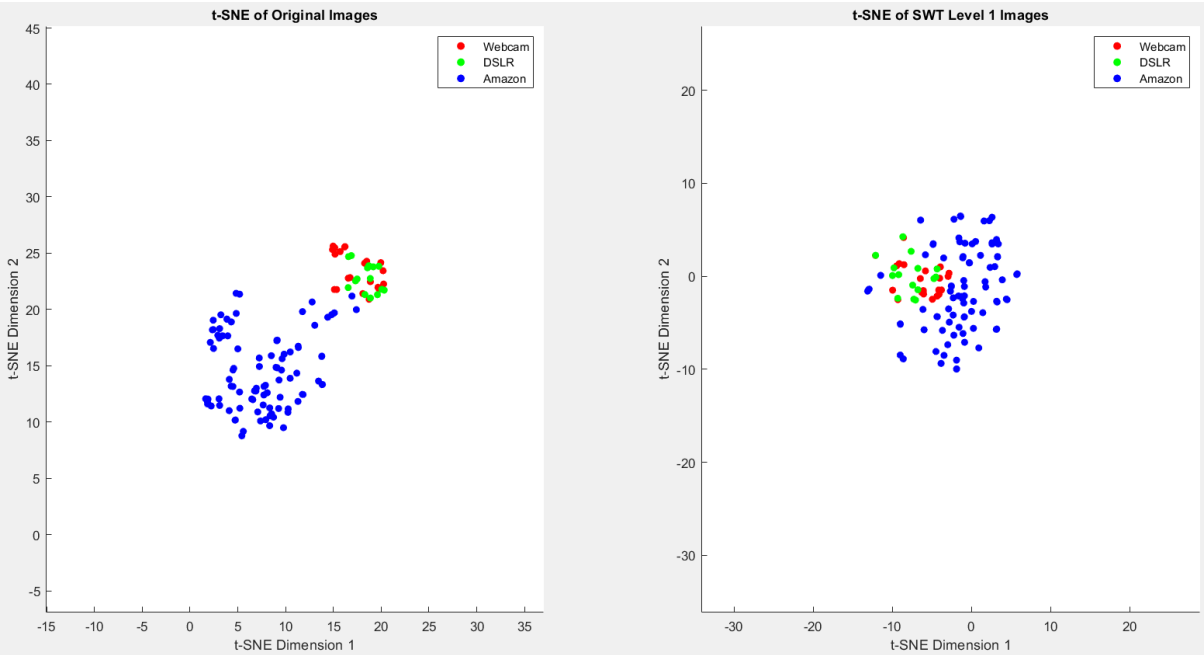

Fig. 26: t-SNE of images from class desktop-computer of all three domains

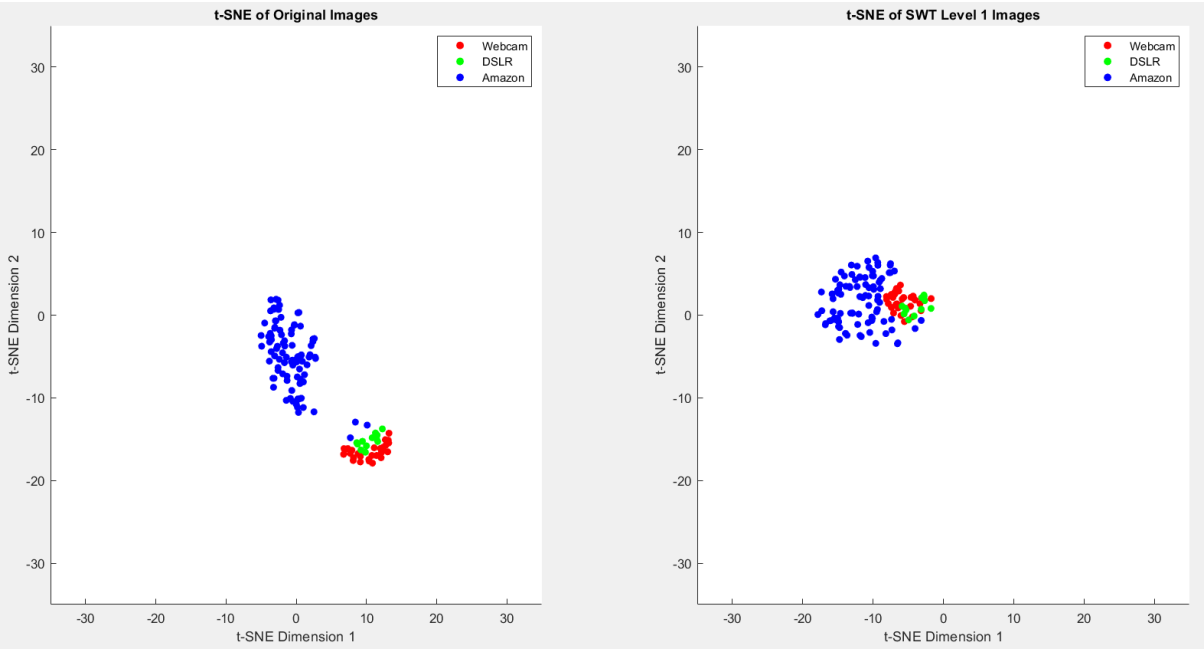

Fig. 27: t-SNE of images from class back-pack of all three domains

6.2 Inference

In the t-SNE of the original images, the points from different domains were separated, highlighting the domain-specific variations. However, after applying SWT and re-plotting the t-SNE, we observed a significant shift: the points from different domains came closer together and got intermixed. This indicates that the SWT effectively reduces domain-specific differences, making it harder to draw a distinct classification boundary. This blending of domains is actually beneficial for cross-domain generalization, as it encourages the model to focus on more robust, invariant features that generalize across domains.
By applying this approach to other different examples of classes, we consistently found that SWT enhances cross-domain feature alignment. This suggests that incorporating wavelets, like SWT, into deep learning architectures can improve generalization and explainability by reducing domain biases and enabling models to learn more transferable, invariant features.

6.2.1 Economy: Reducing Redundant Computation

In our experiment, the application of Scattering Wavelet Transform (SWT) leads to a compression of the data into more relevant, domain-invariant features. This has direct implications for economy:

- **Less Need for Large Datasets:** By focusing on the essential, invariant features across domains, SWT reduces the reliance on extensive labeled datasets from each domain. Models require less data to generalize effectively, which can save both computational time and resources.
- **Reducing Training Time:** As SWT mitigates domain-specific variations early on, deep learning models spend less time learning domain-specific nuances, leading to faster convergence and reduced training cycles.
- **Ease of Dataset Creation:** Not only does this remove the dependence on large data stores, but also significantly decreases the need for normalization and standardization of data. The invariant features from the dataset will possess very similar mathematical properties even without considering the method of obtaining those data points. This means that the SWT also possesses an ability to "self-denoise" the training data.

6.2.2 Explainability: Revealing Domain-Invariant Features

In terms of explainability, our experiment shows that SWT helps in aligning features across domains, reducing domain biases and making the model more interpretable:

- **Clearer Feature Representation:** By applying SWT, we observed that t-SNE plots showed tighter clustering across domains, indicating that the transformation captures

fundamental features that are shared across different environments (like DSLR and webcam images). This makes it easier to interpret what features the model is focusing on, as it no longer needs to differentiate based on domain-specific noise.

- **Enhancing Model Transparency:** Since the wavelet transform is a predefined mathematical operation, it provides a clear, deterministic process for feature extraction, contrasting with the black-box nature of CNNs where learned filters are harder to interpret. This enhances the model's transparency.

References

[1] M. Piduguralla and J. S. Bhatt, "An analytical cnn: Use of wavelets for learning image structures in cross-domain generalization," in *2024 National Conference on Communications (NCC)*, 2024, pp. 1–6.

[2] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[3] M. G. Bergomi, P. Frosini, D. Giorgi, and N. Quercioli, "Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning," *Nature Machine Intelligence*, vol. 1, pp. 423–433, September 2019.

[4] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. USA: Academic Press, Inc., 2008.

[5] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[6] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Analysis of visual domains adaptation in the wild," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, 2010, pp. 1–8.

[7] P. Kumar Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari, "Brain mri image classification for cancer detection using deep wavelet autoencoder-based deep neural network," *IEEE Access*, vol. 7, pp. 46 278–46 287, 2019.