Divergence Regulated Encoder Network for Joint Dimensionality Reduction and Classification

Joshua Peeples, *Student Member, IEEE*, Sarah Walker, *Student Member, IEEE*, Connor Mccurley, *Student Member, IEEE*, Alina Zare, *Senior Member, IEEE*, James Keller, *Life Fellow, IEEE*, and Weihuang Xu, *Student Member, IEEE*

***Abstract*— Feature representation is an important aspect of remote-sensing-based image classification. While deep convolutional neural networks (DCNNs) are able to effectively amalgamate information, large numbers of parameters often make learned features inscrutable and difficult to transfer to alternative models. In order to better represent statistical texture information for remote-sensing image classification, in this letter, we investigate performing joint dimensionality reduction (DR) and classification using a novel histogram neural network. Motivated by a popular DR approach, t-distributed stochastic neighbor embedding (t-SNE), our proposed method incorporates a classification loss computed on samples in a low-dimensional embedding space. We compare the learned sample embeddings against coordinates found by t-SNE in terms of classification accuracy and qualitative assessment. We also explore the use of various divergence measures in the t-SNE objective. The proposed method has several advantages such as readily embedding out-of-sample points and reducing feature dimensionality while retaining class discriminability. Our results show that the proposed approach maintains and/or improves classification performance and reveals characteristics of features produced by neural networks that may be helpful for other applications.**

***Index Terms*— Convolutional neural networks, dimensionality reduction (DR), t-distributed stochastic neighbor embedding (t-SNE).**

I. Introduction

RECENTLY, deep convolutional neural network (DCNN)-based approaches have shown remarkable performance on remote-sensing tasks [1]. State-of-the-art image classification networks often contain thousands or even millions of parameters that need to be learned during training. While it has been argued in the literature that high-dimensionality (data, parameterization, model complexity) is more amenable for classification [2], [3], methods taking advantage of high-dimensionality are often inscrutable. One approach for handling the nuances associated with high-dimensionality is simply to reduce the dimensionality of the data/model.

Dimensionality reduction (DR) has proved to be a critical tool in many remote-sensing applications, such as hyperspectral land-cover classification [4], [5], multi-modality sensor fusion [6], co-registration and alignment [7], redundancy removal, memory usage minimization, and reduction of the effects of the Curse of Dimensionality [8]. The goal of DR can be posed as discovering intrinsic, lower-dimensional features from the data which meet an overarching objective, such as preserving variance, finding compressed representations of data, maintaining global or local structure, or promoting class discriminability in the embedding space [8]–[10]. Most studies perform classification or regression after applying unsupervised DR. However, it has been shown that there are advantages to learning the low-dimensional representations and classification/regression models simultaneously [11], [12]. Specific to classification, the goal of DR is to discover embedding functions that take data from the input feature space and transform it into a lower-dimensional coordinate system or *latent space*. Ideally, the low-dimensional features capture "useful" properties of the data while enforcing constraints such as topological ordering and class separability [13].

The representation of data is a critical factor in determining machine learning classification performance [9]. For images, one can generally represent information through color, shape, and texture features [14]. As shown in the literature [15], texture features often serve as the most powerful descriptor of the three. As currently constructed, DCNNs cannot directly model statistical texture information without the use of extra layers and parameters. To capture statistical texture features in DCNNs, *histogram layer* networks [16] were introduced to characterize the distribution of features in the model. We hypothesize that incorporating statistical texture features into our proposed approach will assist in providing meaningful representations of the data for classification.

Based on these motivations, we introduce a neural classification approach that inherently learns compressed feature representations optimized for class discriminability. The contributions of this work are summarized as the following.

1) We propose a neural classification scheme that finds discriminative, low-dimensional representations of the input data. The network learns from a t-distributed stochastic neighbor embedding (t-SNE)-based objective,

Manuscript received September 12, 2021; revised December 23, 2021 and February 1, 2022; accepted March 1, 2022. Date of publication March 2, 2022; date of current version March 24, 2022. This work was supported in part by the National Science Foundation Graduate Research Fellowship under Grant DGE-1842473 and in part by the Office of Naval Research under Grant N00014-16-1-2323. *(Corresponding author: Joshua Peeples.)*

Joshua Peeples, Sarah Walker, Connor Mccurley, Alina Zare, and Weihuang Xu are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32608 USA (e-mail: jpeeples@ufl.edu).

James Keller is with the Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211 USA.

This article has supplementary downloadable material available at https://doi.org/10.1109/LGRS.2022.3156532, provided by the authors.

Digital Object Identifier 10.1109/LGRS.2022.3156532

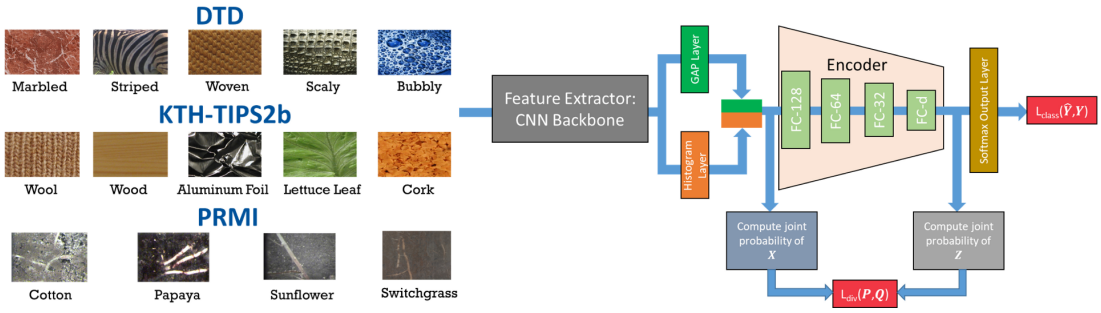


Fig. 1. Architecture for the proposed DRENs.

yet outperforms the data representations found by classical t-SNE in terms of classification accuracy.
2) The proposed method is constructed from a histogram layer network backbone which extracts highly descriptive texture features.
3) The proposed approach can readily embed out-of-sample points, allowing it to be applied universally to other image classification datasets.

II. Methods

A. Proposed Approach

In order to jointly perform classification and DR, our objective function is comprised of two terms: classification loss, L_{class}, and a divergence measure, L_{div}

$$L_{\text{total}} = (1-\lambda)L_{\text{class}}\left(\hat{\mathbf{Y}}, \mathbf{Y}\right) + \lambda L_{\text{div}}(\mathbf{P}, \mathbf{Q}) \tag{1}$$

where $\mathbf{P} \in \mathbb{R}^{N\times N}$ is the joint probability distribution on the high-dimensional input features $\mathbf{X} \in \mathbb{R}^{N\times D}$ for N samples in a mini-batch, $\mathbf{Q} \in \mathbb{R}^{N\times N}$ is the joint probability distribution on the lower-dimensional embeddings $\mathbf{Z} \in \mathbb{R}^{N\times d}$, $\hat{\mathbf{Y}} \in \mathbb{R}^{N\times C}$ are the sample predictions for C classes, and $\mathbf{Y} \in \mathbb{R}^{N\times C}$ are the true class labels. We add a weight, λ, to leverage the contribution of each term to the overall objective. We define L_{class} as the popular cross-entropy loss

$$L_{\text{class}}\left(\hat{\mathbf{Y}}, \mathbf{Y}\right) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log \hat{y}_{ic}. \tag{2}$$

For t-SNE, Kullback–Leibler (KL) divergence is typically used to minimize the divergence between the higher- and lower-dimensional joint probability distributions

$$L_{\text{div}}(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}\log\frac{p_{ij}}{q_{ij}}. \tag{3}$$

Following the standard t-SNE [17] implementation, we used a normalized radial basis kernel and Student's t-distribution with a single degree of freedom to estimate the neighbor probability matrices $\mathbf{P}$ and $\mathbf{Q}$ in the high- and low-dimensional feature spaces, respectively. Our approach is agnostic to the selection of a divergence measure. Hence, we used other divergence measures such as Wassertein-1 (i.e., earth mover's distance) and Renyi's divergence with $\alpha = 0.5$. There was not a statistically significant difference in the classification performance of our model. Therefore, the results presented in this work are representative of using KL divergence.

B. Implementation

For our divergence regulated encoder network (DREN), we used two pretrained models as the backbones, ResNet18 and ResNet50 for Textures under varying Illumination, Pose and Scale (KTH-TIPS-2b) [18] and Describable Texture Dataset (DTD) [19], respectively. We added a histogram layer after the last convolutional layer for each of the pretrained models. From these models, we extracted the high-dimensional feature vectors, $\mathbf{X}$, for each sample. For the baseline and histogram models used for KTH-TIPS-2b and DTD, D was 512, 1024, 2048, and 4096, respectively. Our encoder was composed of four fully connected layers of size 128, 64, 32, and d. We used ReLU activation functions for the first three layers and passed the d-D embedding to a softmax output layer for classification.

The embeddings, $\mathbf{Z}$, were used to compute L_{div} and to update the weights of the model via backpropagation. To create a fair comparison to t-SNE applied to the high-dimensional features, we only updated the lower-dimensional embeddings. Therefore, $\mathbf{P}$ was removed from the computational graph such that only the gradient information from $\mathbf{Q}$ was used. The overall structure of DREN is shown in Fig. 1 and the code for our work is publicly available: https://github.com/GatorSense/DREN.

III. Experimental Setup

Our experiments consisted of two main components: encoder performance and comparisons to t-SNE. For our encoder experiments, we investigated the selection of hyperparameters λ and d for the proposed DREN models. We used two texture/material datasets: DTD and KTH-TIPS-2b. DTD [19] is a collection of textural images annotated with human-centric attributes. The texture database consists of 5640 images, organized according to 47 categories. The KTH-TIPS-2b [18] dataset is defined by images of multiple materials under varying pose, scale, and illumination. The database contains images of 11 distinct textures. The typical train/test splits for DTD (train on 80 images per class and test on 40 images per class) and KTH-TIPS-2b (train on three samples, test on one sample) were used for our experiments [15]. We held

TABLE I

ENCODER EXPERIMENTS PERFORMANCE OF THE DIFFERENT VARIATIONS OF DREN FOR KTH-TIPS-2B. THE BEST AVERAGE RESULT FOR EACH MODEL d, AND λ IS BOLDED. THE PERFORMANCE AND DIMENSIONALITY WITHOUT AN ENCODER ARE SHOWN IN THE HEADINGS. FOR CLARITY, THE STANDARD DEVIATIONS ARE NOT SHOWN

	Resnet18 ($D = 512$, 78.77 $\pm$ 2.86)					HistRes18 ($D = 1024$, 77.96 $\pm$ 2.12)				
Weight (λ)	$d = 2$	$d = 3$	$d = 4$	$d = 8$	$d = 16$	$d = 2$	$d = 3$	$d = 4$	$d = 8$	$d = 16$
$\lambda = 0.0$	75.11	77.55	78.64	78.11	**80.37**	74.94	75.84	**80.11**	78.09	76.87
$\lambda = 0.1$	73.38	76.28	76.35	78.09	75.53	**75.74**	77.74	79.08	75.08	74.98
$\lambda = 0.2$	**76.91**	78.47	78.39	76.81	78.72	74.28	77.06	78.05	78.24	**78.96**
$\lambda = 0.3$	76.41	**78.24**	**78.79**	**79.50**	77.19	75.06	76.22	78.58	77.19	77.50
$\lambda = 0.4$	71.15	74.62	75.82	77.15	73.61	73.78	78.16	77.00	75.93	77.61
$\lambda = 0.5$	70.98	76.79	77.46	77.27	78.39	73.19	**80.07**	77.74	75.27	78.22
$\lambda = 0.6$	71.97	75.53	75.65	**78.43**	77.63	74.01	76.96	76.49	**79.82**	78.45
$\lambda = 0.7$	73.46	73.40	76.41	78.14	78.77	71.28	73.59	79.99	77.21	76.87
$\lambda = 0.8$	69.82	73.78	75.13	76.62	77.50	73.99	75.04	73.61	71.55	74.18
$\lambda = 0.9$	70.29	70.88	72.29	76.77	73.88	71.51	74.24	73.55	78.37	75.63
$\lambda = 1.0$	17.45	14.23	9.41	16.14	17.34	15.89	13.87	18.69	16.16	15.49

10% of the training data for validation and applied each model to the holdout test set. We performed four and ten runs of data splits for KTH-TIPS-2b and DTD, respectively. The average classification accuracy is reported across the different folds. We followed a similar training procedure and data augmentation from previous works [16], except Adam optimization was used. We also only updated newly added layers (i.e., histogram layers, encoder, output layer) and kept the pretrained weights fixed. The number of bins used in all histogram models was 16.

The weight in the objective term, λ, was varied from 0 to 1 in steps of 0.1. The embedding dimension, d, was set to 2, 3, 4, 8, and 16. After our encoder experiments, we compared the embeddings learned by DREN to t-SNE. In order to embed test samples for t-SNE, we used an approach similar to Locally Linear Embedding [20], which assumes that a sample can be represented as a linear combination of its nearest neighbors. Thus, low-dimensional coordinates of unseen test points were found as constrained combinations of their nearest neighbors in the training set. We evaluated each embedding 1) qualitatively and 2) quantitatively by reporting the classification accuracy of a k-nearest neighbors (K-NN) classifier on the test data ($K = 3$). Lastly, we further validated the approach by investigating various convolutional neural networks (CNN) backbone architectures. Additionally, we also show the effectiveness of the approach on a remote-sensing dataset, Plant Root Minirhizotron Imagery (PRMI) [21]. We performed species classification on a subset of the data for a total of 4144 images and four classes. The published training, validation, and test splits were used across five experimental runs.

IV. RESULTS AND DISCUSSION

A. Encoder Experiments

For the two hyperparameters λ and d, the performance of the DREN models depends more on the selection of the embedding dimension. Ideally, as the dimensionality of the feature space becomes larger, inter-class separability should increase and intra-class variations should decrease, leading to improved performance (as shown in our experiments for the DREN models). However, our DREN models achieved statistically comparable performance to the baseline models without an encoder at significantly lower dimensions (except for DREN with ResNet50 for DTD). Our method is also stable to the selection of λ except for when only the L_{div} is considered ($\lambda = 1$). In this case, the model is an unsupervised approach and will only learn embeddings that minimize the divergence between higher and lower dimensions. In our implementation, the error was not backpropagated through the output layer once $\lambda = 1$. In this instance, the output layer will not be updated, leading to poor classification performance. We also analyze the convergence of the proposed method across the various hyperparameters in the supplemental material.

In Table I, the performance of the DREN models with and without the histogram layers for KTH-TIPS-2b was comparable. KTH-TIPS-2b primarily consisted of images collected in a controlled environment [15]. As a result, the local, texture features extracted by the histogram layers may not add as much information for the classification. However, the DREN models with the histogram layers performed better on DTD across the different combinations of hyperparameters as shown in Table II. DTD contained images collected "in-the-wild" and also has more classes (47) than KTH-TIPS-2b (11). The embeddings learned by the DREN histogram models retained useful information from the texture features to achieve comparable performance to the base CNN models.

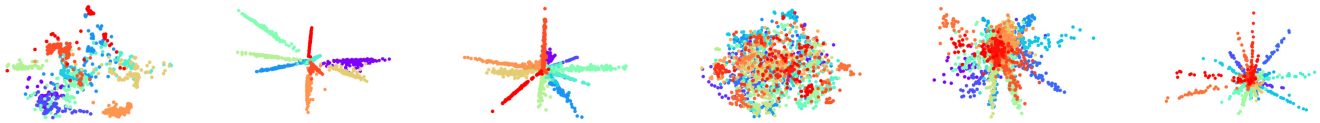
B. t-SNE Comparisons

In Table III, the test performance for the K-NN classifiers showed the utility of the DREN embeddings for not only visualization, but also classification. Our DREN models also have an advantage over t-SNE in that new images can be embedded without a separate out-of-sample approach. In Fig. 2, the t-SNE embedding of the test points is not as visually compact and separable as the DREN models. The test sample embeddings produced from the DREN models also seem to have "angular" feature distributions. As noted in other works [22], models trained using cross-entropy appear to learn these unique, intrinsic feature coordinates in comparison to the t-SNE embeddings that are learned by only using KL divergence. For the DREN histogram models, the "angular" embeddings may also be produced as a result of normalization as the features are also centered around 0. The features

TABLE II

ENCODER EXPERIMENTS PERFORMANCE OF THE DIFFERENT VARIATIONS OF DREN FOR DTD. THE BEST AVERAGE RESULT FOR EACH MODEL d, AND λ IS BOLDED. THE PERFORMANCE AND DIMENSIONALITY WITHOUT AN ENCODER ARE SHOWN IN THE HEADINGS. FOR CLARITY, THE STANDARD DEVIATIONS ARE NOT SHOWN

	Resnet50 ($D = 2048$, 62.79 ± 0.94)					HistRes50 ($D = 4096$, 62.78 ± 1.12)				
Weight (λ)	$d = 2$	$d = 3$	$d = 4$	$d = 8$	$d = 16$	$d = 2$	$d = 3$	$d = 4$	$d = 8$	$d = 16$
$\lambda = 0.0$	28.25	**45.49**	49.74	55.46	62.64	**40.07**	53.06	55.46	**59.74**	62.78
$\lambda = 0.1$	30.34	44.77	50.57	54.96	62.38	39.72	53.49	56.21	59.10	62.65
$\lambda = 0.2$	**30.55**	44.23	**51.27**	55.50	62.49	38.77	**53.08**	56.41	59.11	62.12
$\lambda = 0.3$	30.39	45.10	50.47	55.70	63.01	36.54	52.97	**56.60**	59.01	61.61
$\lambda = 0.4$	29.03	43.44	50.77	56.16	62.58	35.42	50.61	56.05	59.05	61.93
$\lambda = 0.5$	29.41	42.80	50.34	56.16	**63.16**	35.56	48.54	56.07	58.92	62.37
$\lambda = 0.6$	26.93	42.83	50.17	56.12	62.49	34.51	48.45	53.87	59.30	62.83
$\lambda = 0.7$	26.24	40.39	49.57	**56.24**	62.57	32.14	46.30	52.8	59.53	62.03
$\lambda = 0.8$	23.31	38.89	44.66	56.09	62.43	28.19	43.47	49.48	59.54	62.41
$\lambda = 0.9$	18.04	33.60	41.19	54.32	62.51	23.17	36.51	43.16	57.85	**62.96**
$\lambda = 1.0$	2.97	3.38	2.60	3.43	2.28	2.89	3.49	3.29	2.23	2.92



(a) (b) (c) (d) (e) (f)

Fig. 2. Two-dimensional embeddings of test images from t-SNE out-of-sample approach and different variations of DREN for (a)–(c) KTH-TIPS-2b and (d)–(f) DTD. The colors represent different classes from each dataset. In the 2-D space, the HistRes features are more separable according to an angular measure and are also much more compact, that is, each class mostly varies across a single principal vector, and the principal vectors are well separated by an angular margin.

TABLE III

TEST CLASSIFICATION PERFORMANCE USING K-NN "TRAINED" ON EMBEDDINGS FROM T-SNE AND DREN MODELS. FOR THE DREN MODELS, THE WEIGHT CORRESPONDING TO THE BEST AVERAGE PERFORMANCE IS REPORTED. THE BEST AVERAGE RESULT IS BOLDED

Dataset	Embedding Dimension (d)	t-SNE	DREN ResNet	DREN HistRes
KTH-TIPS-2b	2	76.70±2.88	**76.79± 3.18** ($\lambda = .2$)	74.60±3.47 ($\lambda = .1$)
	3	76.91±3.40	78.18±3.40 ($\lambda = 0$)	**79.23±3.63** ($\lambda = .5$)
DTD	2	**43.37±1.01**	29.35±1.22 ($\lambda = .2$)	40.44±1.64 ($\lambda = 0$)
	3	43.55±1.22	44.28±1.57 ($\lambda = 0$)	**53.02±1.24** ($\lambda = .1$)

TABLE IV

TEST CLASSIFICATION PERFORMANCE WITH DIFFERENT BACKBONE ARCHITECTURES FOR THE DREN MODELS. FOR THE DREN MODELS (ENCODER USED), WE USE EQUAL WEIGHTING FOR EACH TERM IN THE OBJECTIVE ($\lambda = 0.5$) AND THE EMBEDDING DIMENSION IS $d = 16$. THE BEST AVERAGE RESULT FOR EACH MODEL IS BOLDED. FOR CLARITY, THE STANDARD DEVIATIONS ARE NOT SHOWN. THE DIMENSIONALITIES (D) OF THE BASE MODELS ARE SHOWN IN THE HEADINGS

Dataset	Encoder	Histogram Layer	ResNext50 ($D = 2048$)	WideResNet50 ($D = 2048$)	DenseNet121 ($D = 1024$)	EfficientNet ($D = 1280$)	RegNet ($D = 400$)
KTH-TIPS-2b	No	No	57.49	74.03	**80.49**	78.64	75.25
		Yes	72.29	65.80	**78.45**	74.83	74.54
	Yes	No	77.02	73.84	**79.52**	76.94	76.68
		Yes	76.03	73.72	**81.25**	76.66	74.56
DTD	No	No	26.22	**60.18**	46.30	49.53	47.62
		Yes	46.72	**57.67**	55.19	45.76	54.58
	Yes	No	53.69	53.32	**57.48**	52.88	53.77
		Yes	56.86	56.06	**60.91**	51.37	55.37
PRMI	No	No	99.05	97.90	**99.18**	96.90	98.25
		Yes	97.90	95.98	98.72	97.72	**98.83**
	Yes	No	99.20	97.95	**99.47**	98.10	99.17
		Yes	98.37	97.58	**99.35**	97.87	98.63

captured by the histogram layer binning function are naturally between 0 and 1, while the features from the global average pooling layer are followed by batch normalization [16].

C. Backbone Experiments

The proposed DREN model improved the average test classification accuracy across various backbone architectures for the baseline and histogram layer models as shown in Table IV. We also observe that the features extracted from the histogram layer also led to the best performance when coupled with the divergence regulation term for the KTH and DTD datasets. The structural changes of the roots are the most prominent features in the PRMI dataset, so the statistical information of the histogram-based models did not impact

performance much in comparison to the base model. The best model across all datasets was the DenseNet model. This model has a unique feature to encourage feature reuse [23] and this was more effective with the divergence regulation of the DREN model. As shown from the experiments, the DREN model is also fairly robust to the CNN backbone architecture.

V. Conclusion

In this work, we presented DREN models for joint DR and classification. The proposed approach learns discriminative features at lower dimensions that can be used for different tasks. The approach has several advantages and is a general framework that can use various classification/divergence measures and deep learning models.

Acknowledgment

The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] C. Shi, X. Zhang, J. Sun, and L. Wang, "Remote sensing scene image classification based on self-compensating convolution neural network," *Remote Sens.*, vol. 14, no. 3, p. 545, Jan. 2022.

[2] A. N. Gorban and I. Y. Tyukin, "Blessing of dimensionality: Mathematical foundations of the statistical physics of data," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 376, no. 2118, Apr. 2018, Art. no. 20170237.

[3] W. Ronny Huang *et al.*, "Understanding generalization through visualizations," 2019, *arXiv:1906.03291*.

[4] Y. Duan, H. Huang, and Y. Tang, "Local constraint-based sparse manifold hypergraph learning for dimensionality reduction of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 613–628, Jan. 2021.

[5] F. Luo, Z. Zou, J. Liu, and Z. Lin, "Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.

[6] D. Hong, N. Yokoya, J. Chanussot, and X. Xiang Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," 2018, *arXiv:1812.11501*.

[7] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, Jan. 2019.

[8] L. van der Maaten, E. Postma, and H. Herik, "Dimensionality reduction: A comparative," *J. Mach. Learn. Res.*, vol. 10, p. 13, Oct. 2007.

[9] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," 2012, *arXiv:1206.5538*.

[10] N. Thorstensen, "Manifold learning and applications to shape and image processing," Ph.D. dissertation, Ecole Nationale des Ponts et Chaussees, Paris, France, Nov. 2009.

[11] G. Chao, Y. Luo, and W. Ding, "Recent advances in supervised dimension reduction: A survey," *Mach. Learn. Knowl. Extraction*, vol. 1, no. 1, pp. 341–358, 2019.

[12] I. Rish, G. Grabarnik, G. A. Cecchi, F. Pereira, and G. J. Gordon, "Closed-form supervised dimensionality reduction with generalized linear models," in *Proc. 25th Int. Conf. Mach. Learn.*, Jan. 2008, pp. 832–839.

[13] E. Vural and C. Guillemot, "A study of the classification of low-dimensional data with supervised manifold learning," 2015, *arXiv:1507.05880*.

[14] P. Hiremath and J. Pujari, "Content based image retrieval based on color, texture and shape features using image and its complement," *Int. J. Comput. Sci. Secur.*, vol. 1, no. 4, pp. 25–35, 2007.

[15] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen, "From BoW to CNN: Two decades of texture representation for texture classification," *Int. J. Comput. Vis.*, vol. 127, no. 1, pp. 74–109, Jan. 2018.

[16] J. Peeples, W. Xu, and A. Zare, "Histogram layers for texture analysis," *IEEE Trans. Artif. Intell.*, early access, Dec. 15, 2021, doi: 10.1109/TAI.2021.3135804.

[17] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[18] P. Mallikarjuna, A. T. Targhi, M. Fritz, E. Hayman, B. Caputo, and J.-O. Eklundh, "The KTH-TIPS2 database," Comput. Vis. Active Perception Lab., Stockholm, Sweden, Tech. Rep. 11, 2006, vol. 11.

[19] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3606–3613.

[20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[21] W. Xu *et al.*, "PRMI: A dataset of minirhizotron images for diverse plant root study," in *Proc. AAAI Conf. Artif. Intell. AI Agricult. Food Syst. (AIAFS) Workshops*, 2022, pp. 838–839.

[22] H. Choi, A. Som, and P. Turaga, "AMC-loss: Angular margin contrastive loss for improved explainability in image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 838–839.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.