

GROUSE: A Task and Model Agnostic Wavelet-Driven Framework for Medical Imaging

Eleonora Grassucci , *Graduate Student Member, IEEE*, Luigi Sigillo , *Graduate Student Member, IEEE*, Aurelio Uncini , *Senior Member, IEEE*, and Danilo Comminiello , *Senior Member, IEEE*

Abstract—In recent years, deep learning has permeated the field of medical image analysis gaining increasing attention from clinicians. However, medical images always require specific preprocessing that often includes downscaling due to computational constraints. This may cause a crucial loss of information magnified by the fact that the region of interest is usually a tiny portion of the image. To overcome these limitations, we propose GROUSE, a novel and generalizable framework that produces salient features from medical images by *grouping* and *selecting* frequency sub-bands that provide approximations and fine-grained details useful for building a more complete input representation. The framework provides the most enlightening set of bands by learning their statistical dependency to avoid redundancy and by scoring their informativeness to provide meaningful data. This set of representative features can be fed as input to any neural model, replacing the conventional image input. Our method is task- and model-agnostic, thus it can be generalized to any medical image benchmark, as we extensively demonstrate with different tasks, datasets, and model domains. We show that the proposed framework enhances model performance in every test we conduct without requiring ad-hoc preprocessing or network adjustments.

Index Terms—Generalizable deep learning, medical image analysis, mutual information, quaternion wavelet transform.

I. INTRODUCTION

DEEP learning methods have led to a breakthrough in medical image analysis, learning to act as clinicians if carefully designed and trained on representative data [1], with promising results in a variety of tasks [2], [3], [4]. However, most of these approaches cannot be directly applied to medical images, as they require careful and ad-hoc preprocessing steps, including resizing and downscaling due to computational constraints. Unfortunately, this procedure may compromise image

Manuscript received 7 June 2023; revised 11 August 2023; accepted 18 September 2023. Date of publication 2 October 2023; date of current version 13 October 2023. The work of Aurelio Uncini was supported in part by Sapienza University of Rome under Grant RM120172AC5A564 C, and in part by PNRR MUR under Grant PE0000013-FAIR. This work was supported in part by Regione Lazio under Grant 21027NP000000136, and in part by PNRR MUR Project “D34Health - Spoke 1” under Grant PNC1221852F49EDDD. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yue Deng. (Corresponding author: Eleonora Grassucci.)

Eleonora Grassucci, Aurelio Uncini, and Danilo Comminiello are with the Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, 00185 Rome, Italy (e-mail: eleonora.grassucci@uniroma1.it; aurelio.uncini@uniroma1.it; danilo.comminiello@uniroma1.it).

Luigi Sigillo is with the Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, 00185 Rome, Italy, and also with the Leonardo Labs, 00195 Rome, Italy (e-mail: luigi.sigillo@uniroma1.it).

Digital Object Identifier 10.1109/LSP.2023.3321554

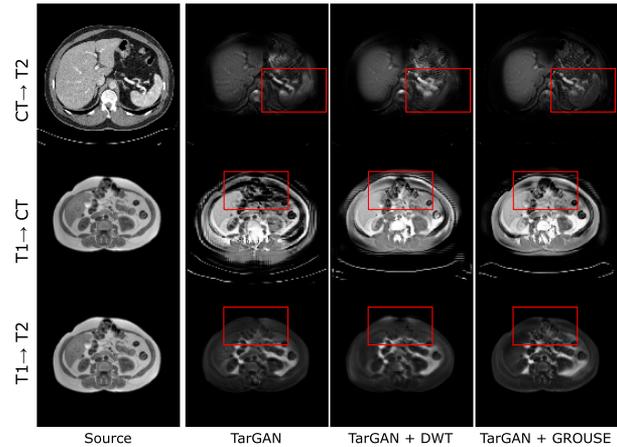


Fig. 1. Random samples of TarGAN translations on the CHAOS dataset, where the first column refers to the source image. The TarGAN produces blurred translations, while when the input is first processed with GROUSE, the translation is far better, as highlighted by the red boxes.

quality losing fine-grained details crucial in this kind of data. Additionally, the region of interest (ROI) in a medical image is usually a tiny portion of the whole acquisition and the model may suffer before effectively focusing on the ROI. Therefore, the prior processing of such data is essential [5], together with the task of prior extracting salient features from samples to help network training.

To alleviate these issues, a literature branch proposes to fuse data modalities in a single sample with more informative contents [6], [7]. Among these methods, wavelet functions have reached promising results, as they process the input at multiple resolutions and handle its low and high frequencies [8], [9], [10]. This transform has been recently widely adopted in deep learning frameworks to distill information, help generation, and analyze or enhance data at a multiresolution scale for vision [11], [12], [13], [14], [15]. However, due to the lack of conventional wavelet shift-invariance property, latest works involve the quaternion wavelet transform (QWT) that provides 16 sub-bands being also shift-invariant [16], [17], [18]. Recently, the adoption of quaternion algebra for neural networks applications is proliferating due to the ability of these methods to preserve the multidimensional structure of the input [19], [20], [21], [22].

In this letter, we propose GROUSE, a novel task- and model-agnostic framework for medical neural imaging. The first step of the framework consists in extracting bands at different frequencies from an input image to produce a set of low-

and high-frequency features to be fed as input to any neural network. Low-frequency bands provide an approximation of image content, while high-frequency ones give details useful for fine-grained tasks such as reconstruction and generation. Nevertheless, passing all the extracted features to the model would be counter-productive as the bands may contain redundant information. To this end, the second step of our framework selects the most informative QWT sub-bands by discovering the statistical dependence among the bands, clustering them according to their mutual information score, and finally selecting the most informative from each group. At convergence, the bands from the same group will be statistically dependent on each other while independent of the bands in other groups, thus ensuring that selected features provide independent information to the neural network, and avoiding redundancy.

One of the most important properties of our *grouping and selection*-based approach is that it is task-agnostic, as it can be applied to any dataset and for different tasks, without loss of generalization. It is also model-agnostic since it can be applied to any neural network model. Another significant property of our approach is its generalizability. Indeed, GROUSE can be involved to enhance the performance of real as well as quaternion-valued neural networks, generalizing its adoption also to grayscale and single-channel images. To prove the generalizability of our method, we conduct an extensive experimental evaluation of multiple tasks involving reconstruction, segmentation, and image modality translation on different benchmarks such as grayscale images, RGB ones, and multimodal datasets. Our method improves the performance of the models in every test we perform, proving that the QWT properly extracts salient features from the medical sample and that the proposed algorithm selects the most informative set of sub-bands to use as input.

The rest of the letter is organized as follows. Section II provides quaternion wavelet fundamentals, Section III introduces the selection algorithm, and Section IV reports the experimental evaluation, while we draw conclusions in Section V.

II. QUATERNION WAVELET TRANSFORM FOR MEDICAL IMAGING

Wavelet transforms have been widely used in literature for medical imaging thanks to their properties of extracting information at different frequencies and with multiple resolutions (e.g., [8], [9]).

Quaternions belong to the class of Clifford algebras over \mathbb{R} . A quaternion number is represented by a real-valued component and three imaginary units, i.e. $h = h_0 + h_1\hat{i}_1 + h_2\hat{i}_2 + h_3\hat{i}_3$.

The imaginary units involved in quaternion numbers comply with the properties $\hat{i}_1^2 = \hat{i}_2^2 = \hat{i}_3^2 = -1$; $\hat{i}_1 \times \hat{i}_2 = \hat{i}_3$, $\hat{i}_2 \times \hat{i}_3 = \hat{i}_1$, $\hat{i}_3 \times \hat{i}_1 = \hat{i}_2$, leading to the non-commutativity of the vector product in this domain. Quaternions are widely adopted for color images, due to their ability to encapsulate each channel information in an imaginary component. This is not possible in medical applications where images are usually grayscale. However, wavelet transform allows for filling this gap.

Wavelets analyze data at different resolutions and trim it into different frequency components. They are particularly suitable for analyzing physical data whose signals contain discontinuities and sharp spikes. The discrete wavelet transform (DWT) on a

2D input provides a scaling function $\phi_x\phi_y$ and three sub-band wavelets underlying diagonal $\phi_x\phi_y$, horizontal $\psi_x\phi_y$, and vertical details $\phi_x\psi_y$. However, due to the lack of phase information, the DWT is not shift-invariant, therefore small shifts in image content can affect the magnitude of the wavelet coefficients. However, we can arrange the four output components of a DWT in a quaternion number so as to leverage its magnitude-phase representation, which may alleviate the shift-invariance issue and provide the missing information.

To overcome DWT limitations, we can instead use the quaternion wavelet transform (QWT) that involves a real-valued DWT and its three Hilbert transforms along the x , y , and xy axis, being approximately shift-invariant [23]. The QWT cuts up data into four different quaternion sub-bands, each with four real-valued coefficients. At the end of the process, the QWT produces 16 real sub-bands comprising 4 low-frequency ones and 12 high-frequency ones, resulting in the following formulation:

$$\begin{aligned}\phi_h &= \phi_{g,x}\phi_{g,y} + \phi_{f,x}\phi_{g,y}\hat{i}_1 + \phi_{g,x}\phi_{f,y}\hat{i}_2 + \phi_{f,x}\phi_{f,y}\hat{i}_3 \\ \psi_h^V &= \psi_{g,x}\phi_{g,y} + \psi_{f,x}\phi_{g,y}\hat{i}_1 + \psi_{g,x}\phi_{f,y}\hat{i}_2 + \psi_{f,x}\phi_{f,y}\hat{i}_3 \\ \psi_h^H &= \phi_{g,x}\psi_{g,y} + \phi_{f,x}\psi_{g,y}\hat{i}_1 + \phi_{g,x}\psi_{f,y}\hat{i}_2 + \phi_{f,x}\psi_{f,y}\hat{i}_3 \\ \psi_h^D &= \psi_{g,x}\psi_{g,y} + \psi_{f,x}\psi_{g,y}\hat{i}_1 + \psi_{g,x}\psi_{f,y}\hat{i}_2 + \psi_{f,x}\psi_{f,y}\hat{i}_3,\end{aligned}\quad (1)$$

in which g is a filter and f the corresponding Hilbert transform, producing the scaling function ϕ_h that encloses low-frequency coefficients and three wavelets $\psi_h^V, \psi_h^H, \psi_h^D$ with high-frequency details. A common choice for QWTs is considering the Daubechies wavelet with 8 vanishing moments (db8) [24]. From this DWT we can then extract the decomposition low-pass G_L and high-pass filters G_H to which the Hilbert transform is applied in order to obtain the counterpart filters F_L and F_H . The QWT is computed by combining the filters in a couple, and interleaving them with a downsampling operation [24], [25]. We show the QWT sub-bands for a samples image in Fig. 2, where the four low-frequency bands $L_g L_g, L_g L_f, L_f L_g$, and $L_f L_f$ are displayed in the first column and the twelve high-frequency $L_g H_g, L_g H_f, L_f H_g, L_f H_f, H_g L_g, H_g L_f, H_f L_g, H_f L_f, H_g H_g, H_g H_f, H_f H_g$, and $H_f H_f$ in the last three columns.

III. THE PROPOSED GROUSE FRAMEWORK

A common approach in literature for selecting real-valued wavelet features is involving low-frequency sub-bands that have been proven to improve classification accuracy [26]. A similar method proposes to compute the energy for each sub-band and then get the top k features with the highest energy [27]. However, these methods aim at selecting separately the more informative or discriminative sub-bands. Indeed, while low-frequency bands may be the most informative and obtain the highest energy values, they may miss crucial details and high-frequency characteristics suitable for proper signal analysis and synthesis. Moreover, previous approaches implicitly assume independence among sub-bands that does not hold in practice. For this reason, when choosing the best sub-bands to feed as input to the model, the selection method has to take into account both the informativeness of the single sub-band and the total information

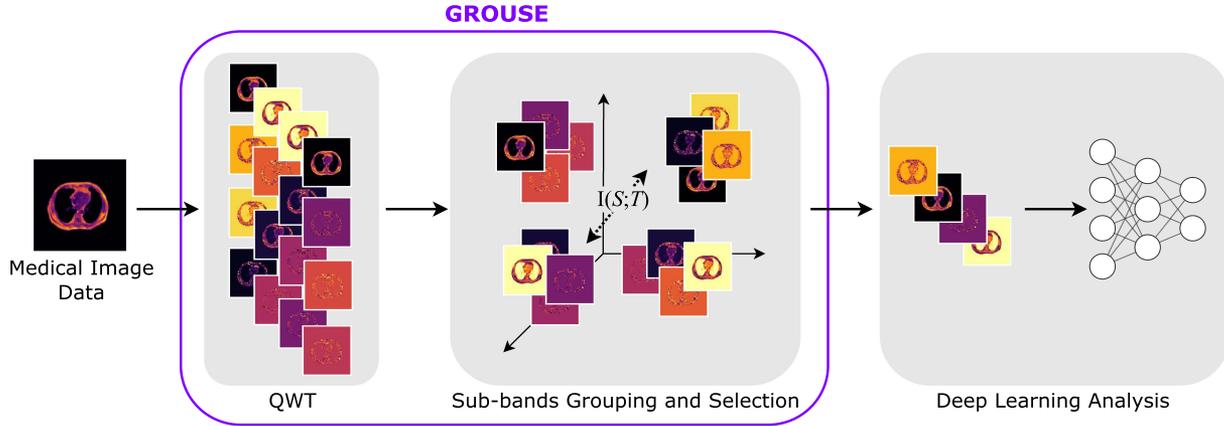


Fig. 2. Proposed GROUSE framework. Medical images are processed with quaternion wavelet transform, grouped according to the mutual information, and then the highest-energy sub-bands are selected to be fed as input to any neural model for analysis.

brought by the k sub-bands combined together. To this end, we take inspiration from an approach that builds clusters for real DWT tree sub-bands considering the dependence among them and then perform an evaluation of individual sub-bands in order to improve the accuracy of texture classification [28]. We propose a quaternion wavelet transform sub-band GROUPing and SELECTION (GROUSE) approach that completely operates in a quaternion fashion while being easy to be incorporated into any real-valued analysis. The algorithm first discovers the structure of the statistical dependence among QWT sub-bands, split them into clusters based on the dependence value, then an intra-cluster evaluation of wavelet features is performed by computing the energy of each element, selecting the sub-band with the highest energy as a candidate for the final set. While the original approach [28] computes the DWT sub-band selection on the training set and then involves the discovered sub-bands for texture classification on the test set only, we insert a selection method in the framework that preprocesses each input image, and then with the preprocessed data, we perform model training. This helps the model to take advantage of the QWT extracted features during training and consequently improve the performance at the testing stage. The potential of our approach is that it does not require labels or additional information, so it can be involved both in the training and testing stages, as preferred by the user.

Energy function: To be consistent with the method adapted for sub-band selection [28], in this letter we employ the squared magnitude as the energy function, computed as:

$$E_{i,j}(\mathbf{s}_{i,j}) = \sum_{h=1}^H \sum_{w=1}^W [s_{i,j,(h,w)}]^2 \quad (2)$$

whereby H and W are the height and width of the sub-band \mathbf{s}_j of the image \mathbf{X}_i . In this way, the energy of a sub-band is regarded as a second-order polynomial in terms of the pixel values of the considered image.

Statistical dependency: The optimal result of the statistical dependency discovering analysis is that sub-bands from the same cluster are strongly dependent, while sub-bands coming from diverse clusters are independent. If the clustering algorithm converges to this solution, then it is possible to build

a meaningful representation of the input by selecting just the most representative sub-band from each cluster. As a metric for estimating the amount of dependency of a sub-band with respect to a set of features we involve the mutual information (I) [28]. Indeed, the mutual information quantifies how much information the set of features contains about the sub-band, being zero when the two objects are independent. The higher the value of I , the better the estimation of the marginal distribution of the sub-band will be. However, the estimate may be inaccurate in case of high-dimensional distributions such as multiple sub-bands and complex input data such as images thus, for this reason, a many-to-one scalar mapping is involved, assuming that the set of features in a cluster provides information about the distribution S of the selected sub-band \mathbf{s} through a function $T = f(\mathbf{X}_1, \dots, \mathbf{X}_N)$. At this point, the mutual information can be directly computed as $I(S; T)$ and remains open the choice of the function $f(\cdot)$. From the data-processing theorem [29] we know that the true mutual information $I(S; \{\mathbf{X}_1, \dots, \mathbf{X}_N\})$ is the upper bound of our approximation and that the equality holds if and only if T is the sufficient statistics for the set of data $\{\mathbf{X}_1, \dots, \mathbf{X}_N\}$. Therefore, the function $f(\cdot)$ has to be chosen to maximize the estimate $I(S; T)$ while remaining convenient in terms of computation. To this end, involving a linear model of the type $T = f(\mathbf{X}_1, \dots, \mathbf{X}_N) = \sum_{i=1}^N \mathbf{W}_i \mathbf{X}_i$, with $\mathbf{W}_i = 1/N$, has been proved to match the required statistical properties being the unbiased estimate of the mean of \mathbf{X}_i [28], [30]. At this point, the mutual information can be computed as:

$$I(S; T) = \sum_{\mathbf{s}} \sum_{\mathbf{t}} P_{S,T}(\mathbf{s}, \mathbf{t}) \log \frac{P_{S,T}(\mathbf{s}, \mathbf{t})}{P_S(\mathbf{s})P_T(\mathbf{t})}, \quad (3)$$

in which $P_{S,T}(\mathbf{s}, \mathbf{t})$ is the joint probability distribution while $P_S(\mathbf{s})$ and $P_T(\mathbf{t})$ are the marginal probability distributions of the selected sub-band and of the unbiased estimate T . A formal definition of the GROUSE algorithm follows.

Algorithm: Given a dataset of images $\{\mathbf{X}_1, \dots, \mathbf{X}_N\}$, the proposed quaternion wavelet sub-band GROUPing and SELECTION (GROUSE) algorithm can be formalized as:

- 1) Compute the QWT as described in Section II and extract the sixteen sub-bands $\mathbf{s}_{i,1}, \dots, \mathbf{s}_{i,16}$ for each image \mathbf{X}_i .

- 2) For each sub-band $s_{1:N,j}$, with $j = 1, \dots, 16$, estimate the marginal probability density function of the energy $E_j(s_{1:N,j})$ in (2) using a Gaussian kernel.
- 3) Apply the k -Means algorithm involving the mutual information in (3) as distance metric and generate the set of clusters $C = \{C_1, \dots, C_k\}$. In our experiments, to properly fit the input also in QNNs, we select $k = 4$.
- 4) From each cluster in C , select the sub-band with the highest energy value, i.e. the most informative for the given partition.
- 5) Involve the set of selected sub-bands as input to the neural model.

We repeat the algorithm 20 times for more robust results.

IV. EXPERIMENTAL EVALUATION

In this section, we validate our theoretical claims on different datasets and for various tasks. We consider the IXI dataset, which comprises brain magnetic resonance (MR) images (T1 and T2) of dimension 256×256 . Then, we take into account the Kvasir-SEG dataset that is composed of 1000 images 256×256 of colon polyps. Finally, we take into consideration the CHAOS dataset that is composed of CT and MR abdominal scans sliced and resized at 128×128 , resulting in 4144 samples. We run the GROUSE algorithm for each dataset, so to obtain the four most representative sub-bands specific to the given data. For the multimodal CHAOS dataset, we run the sub-band selection algorithm specific for each modality, as CT, T1, and T2 are unbalanced in the dataset and are quite different from each other so the salient features may differ from one modality to another. For the IXI and the Kvasir-SEG datasets GROUSE selects the same bands ($L_g L_g, L_f L_g, H_g H_g, L_f H_f$), while for the CHAOS dataset we have three different samples. In the CT modality the result is $L_g L_g, H_g L_g, H_g H_g, L_f L_g$, while for T1 $L_g L_g, L_f L_g, L_f H_g, H_g L_f$ and for T2 $L_g L_g, L_f L_g, H_g L_f, H_g H_f$. It is worth noting that GROUSE always selects the low-frequency $L_g L_g$ band, as it is the most representative of the input and the one with the highest energy.

We perform tests in multiple scenarios, involving reconstruction, segmentation, and image modality translation tasks. As baselines, we consider a vanilla convolutional autoencoder (AE) for reconstruction, a U-Net [31] and a more advanced U-Net++ [32] for segmentation, and the recent large-scale Target-aware generative adversarial networks (TarGAN) [33] for the image modality translation task. For each model, we involve the same hyperparameters setting and network architecture of the original works. Table I shows the objective scores for the three tasks, where for the reconstruction task we compute the structural similarity index (SSIM), the mean squared error (MSE), and the Fréchet Inception distance (FID), for the segmentation task the dice score coefficient (DSC), the mean intersection over union (mIoU) and the mean absolute error (MAE), while for the translation task, we consider a specific FID proposed for medical images in [33], which rewards images that translate better meaningful features. As it is evident from Table I, models endowed with the proposed framework clearly outperform baseline networks that directly involve grayscale or RGB images, proving that our GROUSE approach can produce meaningful features to be involved in neural network training. Additionally, we show that our framework can be employed in various

TABLE I
RESULTS (AVERAGE SCORES OVER MULTIPLE RUNS) FOR RECONSTRUCTION, SEGMENTATION AND IMAGE MODALITY TRANSLATION

Dataset	Reconstruction	Method	SSIM \uparrow	MSE \downarrow	FID \downarrow
IXI	AE	Grayscale	0.856	0.0040	0.178
		DWT	0.846	0.0044	0.201
		GROUSE	0.923	0.0030	0.122
Dataset	Segmentation	Method	DSC \uparrow	mIoU \uparrow	MAE \downarrow
Kvasir-SEG	U-Net [31]	RGB	0.740	0.614	0.073
		DWT	0.745	0.626	0.075
		GROUSE	0.749	0.631	0.071
	U-Net++ [32]	RGB	0.773	0.638	0.066
		DWT	0.776	0.647	0.066
		GROUSE	0.782	0.655	0.065
Dataset	Translation	Method	FID [33] \downarrow		
CHAOS	TarGAN [33]	Grayscale	0.0955		
		DWT	0.0908		
		GROUSE	0.0897		

Bold values represent the best results for each subset of the experiment.

TABLE II
RECONSTRUCTION RESULTS ON THE IXI DATASET FOR VANILLA AUTOENCODERS DEFINED IN THE QUATERNION DOMAIN

Dataset	Reconstruction	Method	SSIM \uparrow	MSE \downarrow	FID \downarrow
IXI	QAE	Channel repet. [34]	0.900	0.0036	0.092
		DWT	0.665	0.0072	0.084
		GROUSE	0.915	0.0031	0.084

Bold values represent the best results for each subset of the experiment.

medical image analysis and synthesis scenarios that comprise different tasks and datasets without requiring specific operations or modifications to either data or models. A visual comparison of the translation task in Fig. 1 underlines the superiority of the TarGAN that involves features provided by our framework. The bounding boxes prove that our model can correctly translate the ROI of the given image while the original fails. Therefore, the task-agnostic GROUSE can be easily generalized to multimodality datasets, while improving the performance of large-scale models too.

Additionally, we also extend our approach to networks defined in the quaternion domain in order to show how our framework can generalize QNNs adoption in onedimensional problems. Indeed, for such tasks, the common approach is to replicate the single channel four times to fit the desired quaternion dimension, so adding useless and redundant information [34]. We prove that our GROUSE framework enhances the performance of QNNs providing a representative set of sub-bands, as Table II shows for the reconstruction task, where the quaternion autoencoder (QAE) with GROUSE data far exceeds both the DWT and the common channel repetition one [34].

V. CONCLUSION

In this letter, we introduce a generalizable framework that can be integrated into any medical image analysis or synthesis approach. Involving the GROUSE output as input to neural models enhances model performance in a variety of tasks and datasets, without requiring specific preprocessing or modifications to existing networks. As a future work, we leave the investigation of GROUSE on different applications such as remote sensing [35], [36].

REFERENCES

- [1] G. Zhao, Q. Feng, C. Chen, Z. Zhou, and Y. Yu, "Diagnose like a radiologist: Hybrid neuro-probabilistic reasoning for attribute-based medical image diagnosis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7400–7416, Nov. 2022.
- [2] N. K. Tomar et al., "FANet: A feedback attention network for improved biomedical image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 25, 2022, doi: [10.1109/TNNLS.2022.3159394](https://doi.org/10.1109/TNNLS.2022.3159394).
- [3] C. Li, W. Chen, X. Luo, Y. He, and Y. Tan, "Adaptive pseudo labeling for source-free domain adaptation in medical image segmentation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2022, pp. 1091–1095.
- [4] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.
- [5] S. Masoudi et al., "Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research," *J. Med. Imag.*, vol. 8, no. 1, pp. 893–912, 2021.
- [6] W. Tang, L. Wang, and Y. Liu, "Green fluorescent protein and phase contrast image fusion via dual attention residual network," in *Proc. IEEE Int. Conf. Med. Imag. Phys. Eng.*, 2021, pp. 1–6.
- [7] Y. Xu, H. Wang, X. Yin, and L. Tao, "MRI and PET/SPECT image fusion based on adaptive weighted guided image filtering," in *Proc. IEEE 5th Int. Conf. Signal Image Process.*, 2020, pp. 275–280.
- [8] Y. Yang, "Medical image fusion through a new DWT based technique," in *Proc. IEEE 4th Int. Conf. Bioinf. Biomed. Eng.*, 2010, pp. 1–4.
- [9] Y. -Hao Yan, L. Li, and Y. -Zhi Yang, "Image fusion based on principal component analysis in dual-tree complex wavelet transform domain," in *Proc. IEEE Int. Conf. Wavelet Act. Media Technol. Inf. Process.*, 2012, pp. 70–73.
- [10] Z. Tao, T. Wei, and J. Li, "Wavelet multi-level attention capsule network for texture classification," *IEEE Signal Process. Lett.*, vol. 28, pp. 1215–1219, 2021.
- [11] W. Ha, C. Singh, F. Lanusse, S. Upadhyayula, and B. Yu, "Adaptive wavelet distillation from neural networks through interpretations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 20669–20682.
- [12] F. Guth, S. Coste, V. De Bortoli, and S. Mallat, "Wavelet score-based generative modeling," in *Proc. Adv. Neural Info. Process. Sys.*, 2022, pp. 478–491.
- [13] M. X. B. Rodriguez et al., "Deep adaptive wavelet network," in *Proc. CVF Winter Conf. Appl. Comput. Vis.*, 2020, pp. 3100–3108.
- [14] Z. Sun and T. Blu, "A nonlinear steerable complex wavelet decomposition of images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2022, pp. 1690–1694.
- [15] Z. Ma and C. Oh, "A wavelet-based dual-stream network for underwater image enhancement," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2022, pp. 2769–2773.
- [16] Z. Zhancheng, L. Xiaoqing, X. Mengyu, W. Zhiwen, and L. Kai, "Medical image fusion based on quaternion wavelet transform and visibility feature," *J. Algorithms Comput. Technol.*, vol. 14, pp. 1–12, 2020.
- [17] P. Geng, X. Sun, and J. Liu, "Adopting quaternion wavelet transform to fuse multi-modal medical images," *J. Med. Biol. Eng.*, vol. 37, pp. 230–239, 2017.
- [18] C. Li, J. Li, and B. Fu, "Magnitude-phase of quaternion wavelet transform for texture representation using multilevel copula," *IEEE Signal Process. Lett.*, vol. 20, no. 8, pp. 799–802, Aug. 2013.
- [19] A. Muppidi and M. Radfar, "Speech emotion recognition using quaternion convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 6309–6313.
- [20] E. Grassucci, D. Comminiello, and A. Uncini, "A quaternion-valued variational autoencoder," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 3310–3314.
- [21] C. Brignone, G. Mancini, E. Grassucci, A. Uncini, and D. Comminiello, "Efficient sound event localization and detection in the quaternion domain," *IEEE Trans. Circuits Syst. II, Express Brief.*, vol. 69, no. 5, pp. 2453–2457, May 2022.
- [22] E. Grassucci, L. Sigillo, A. Uncini, and D. Comminiello, "Hypercomplex image-to-image translation," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2022, pp. 1–8.
- [23] W. L. Chan, H. Choi, and R. G. Baraniuk, "Quaternion wavelets for image analysis and processing," in *Proc. IEEE Int. Conf. Image Process.*, 2004, pp. 3057–3060.
- [24] L. H. Ngo, N. M. Sirakov, M. Luong, E. Viennet, and T. Le-Tien, "Image classification based on sparse representation in the quaternion wavelet domain," *IEEE Access*, vol. 10, pp. 31548–31560, 2022.
- [25] W. L. Chan, H. Choi, and R. G. Baraniuk, "Coherent multiscale image processing using dual-tree quaternion wavelets," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1069–1082, Jul. 2008.
- [26] L. H. Ngo, M. Luong, N. M. Sirakov, T. Le-Tien, S. Guerif, and E. Viennet, "Sparse representation wavelet based classification," in *Proc. IEEE 25th Int. Conf. Image Process.*, 2018, pp. 2974–2978.
- [27] T. Chang and C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Process.*, vol. 2, no. 4, pp. 429–441, Oct. 1993.
- [28] K. Huang and S. Aviyente, "Wavelet feature selection for image classification," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1709–1720, Sep. 2008.
- [29] T. Cover and J. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.
- [30] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1647–1658, Nov. 2001.
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. IEEE Med. Image Comput. Comput. Assist. Intervention*, 2015, pp. 234–241.
- [32] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "U-net++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.
- [33] C. Junxiao, W. Jia, and L. Rui, "TarGAN: Target-aware generative adversarial networks for multi-modality medical image translation," in *Proc. 24th Int. Conf. Med. Image Comput. Comput. Assist. Intervention*, 2021, pp. 24–33.
- [34] T. Parcollet, M. Morchid, and G. Linares, "Quaternion convolutional neural networks for heterogeneous image processing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Brighton, U.K., 2019, pp. 8514–8518.
- [35] L. Sigillo, E. Grassucci, and D. Comminiello, "StawGAN: Structural-aware generative adversarial networks for infrared image translation," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2023, pp. 1–5.
- [36] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.