QUATERNION-VALUED VARIATIONAL AUTOENCODER

Eleonora Grassucci, Danilo Comminiello, and Aurelio Uncini

Dept. Information Engineering, Electronics and Telecommunications (DIET)
Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy

ABSTRACT

Deep probabilistic generative models have achieved incredible success in many fields of application. Among such models, variational autoencoders (VAEs) have proved their ability in modeling a generative process by learning a latent representation of the input. In this paper, we propose a novel VAE defined in the quaternion domain, which exploits the properties of quaternion algebra to improve performance while significantly reducing the number of parameters required by the network. The success of the proposed quaternion VAE with respect to traditional VAEs relies on the ability to leverage the internal relations between quaternion-valued input features and on the properties of second-order statistics which allow to define the latent variables in the augmented quaternion domain. In order to show the advantages due to such properties, we define a plain convolutional VAE in the quaternion domain and we evaluate it in comparison with its real-valued counterpart on the CelebA face dataset.

Index Terms— Variational Autoencoder, Quaternion Neural Networks, Quaternion Properness, Generative Models, Quaternion Random Vectors

1. INTRODUCTION

Variational autoencoders (VAEs) gained their success due to their ability of modeling a generative process formalized in the framework of probabilistic graphical models with latent variables [1, 2]. VAEs are characterized by a simple and fast sampling and easily accessible networks, which have proliferated their use in various applications. Recently, advanced VAEs have been developed, relying on statistical challenges [3, 4] or on hierarchical architectures [5, 6].

Performing learning operations in the quaternion domain, rather than in the real-valued domain, has proved to provide higher efficiency when dealing with multidimensional data [7–9]. Quaternions are widely used in neural networks since they allow to process groups of features together, thus coding latent inter-dependencies between features with a lower

number of parameters with respect to real-valued neural networks [10–12]. These advantages are due to the properties of quaternion algebras, including the Hamilton product that is used in quaternion convolutions. This has recently paved the way to the development of novel deep quaternion neural networks [10, 12, 13], often tailored to specific applications, including theme identification in telephone conversation [14], 3D sound event detection and localization [15, 16], heterogeneous image processing [17] and speech recognition [18]. Other properties of quaternion algebra that may be exploited in learning processes are related to the second-order statistics. In particular, the concept of properness for quaternion valued random signals and their applications has been widely investigated in the signal processing literature [19–23].

In this paper, we extend the concept of properness to the latent space in order to derive a quaternion-valued VAE (QVAE). Designing VAEs in the quaternion domain may bring several advantages. Neural networks for VAEs should model long-range correlations in data [4, 24] and quaternion convolutional layers may provide additional information by leveraging internal latent relations between input features. Moreover, reducing the number of parameters may benefit the decoder and the marginal log-likelihood, which only depends on the generative network [4].

The contribution of this paper is twofold: i) we define how to perform variational inference in the quaternion domain and ii) propose a plain convolutional QVAE. Moreover, to the best of our knowledge, for the first time augmented quaternion second-order statistics are exploited for the development of a deep learning model.

The paper is organized as follows. In Section 2, we review the main learning operations in the quaternion domain. The proposed QVAE is derived in Section 3 and then evaluated in Section 4. Finally, conclusions are drawn in Section 5.

2. LEARNING IN THE QUATERNION DOMAIN

2.1. Main Properties of Quaternion Algebra

The quaternion domain $\mathbb{H}$ is defined by a four-dimensional associative normed division algebra over real numbers, belonging to the class of Clifford algebras. A quaternion is expressed by one real scalar component and three imaginary ones:

$$q = q_a + q_b\hat{\imath} + q_c\hat{\jmath} + q_d\hat{\kappa} = q_a + \overline{q}. \qquad (1)$$

with $q_a, q_b, q_c, q_d \in \mathbb{R}$. The imaginary units, $\hat{\imath} = (1,0,0)$, $\hat{\jmath} = (0,1,0)$, $\hat{\kappa} = (0,0,1)$, are unit axis vectors and represent an orthonormal basis in $\mathbb{R}^3$. The vector $\overline{q} = q_b\hat{\imath} + q_c\hat{\jmath} + q_d\hat{\kappa}$ represents the imaginary part of the quaternion and it is also known as *pure quaternion*. The imaginary units comply with the following fundamental properties:

$$\hat{\imath}^2 = \hat{\jmath}^2 = \hat{\kappa}^2 = -1, \qquad (2)$$

$$\hat{\imath}\hat{\jmath} = \hat{\imath} \times \hat{\jmath} = \hat{\kappa}, \quad \hat{\jmath}\hat{\kappa} = \hat{\jmath} \times \hat{\kappa} = \hat{\imath}, \quad \hat{\kappa}\hat{\imath} = \hat{\kappa} \times \hat{\imath} = \hat{\jmath}, \quad (3)$$

where "$\times$" denotes the vector product in $\mathbb{R}^3$. The above properties allow to adopt quaternion algebra $\mathbb{H}$ to represent spatial rotations in $\mathbb{R}^4$. Quaternion algebra $\mathbb{H}$ is endowed with the operations of associative multiplication of elements in the algebra and scalar multiplication. The scalar product of two quaternions p and q is defined as $q \cdot p = q_a p_a + q_b p_b + q_c p_c + q_d p_d$. However, quaternions are not commutative under the operation of vector multiplication, thus $\hat{\imath}\hat{\jmath} \neq \hat{\jmath}\hat{\imath}$, and consequently $\hat{\imath}\hat{\jmath} = -\hat{\jmath}\hat{\imath}$, $\hat{\jmath}\hat{\kappa} = -\hat{\kappa}\hat{\jmath}$, $\hat{\kappa}\hat{\imath} = -\hat{\imath}\hat{\kappa}$.

Due to the noncommutative property, the quaternion product, also known as Hamilton product, can be expressed as:

$$
\begin{aligned}
qp &= (q_a + q_b\hat{\imath} + q_c\hat{\jmath} + q_d\hat{\kappa})(p_a + p_b\hat{\imath} + p_c\hat{\jmath} + p_d\hat{\kappa}) \\
&= (q_a p_a - q_b p_b - q_c p_c - q_d p_d) \\
&+ (q_a p_b + q_b p_a + q_c p_d - q_d p_c)\hat{\imath} \\
&+ (q_a p_c - q_b p_d + q_c p_a + q_d p_b)\hat{\jmath} \\
&+ (q_a p_d + q_b p_c - q_c p_b + q_d p_a)\hat{\kappa}.
\end{aligned}
\qquad (4)
$$

We can define the conjugate and the module of a quaternion, respectively, as $q^* = q_a - q_b\hat{\imath} - q_c\hat{\jmath} - q_d\hat{\kappa}$ and $|q| = \sqrt{q_a^2 + q_b^2 + q_c^2 + q_d^2} = |q^*|$. A quaternion can be also written in polar form as $q = |q|(\cos(\theta) + \nu\sin(\theta)) = |q|e^{\nu\theta}$, where $\theta \in \mathbb{R}$ is the argument of the quaternion, $\cos(\theta) = q_a/\|q\|$, $\sin(\theta) = \|\overline{q}\|/\|q\|$ and $\nu = \overline{q}/\|\overline{q}\|$ is a pure unit quaternion. The involution of a quaternion q over a pure unit quaternion ν is $q^\nu = -\nu q\nu$, and it represents a rotation of π in the imaginary plane orthogonal to $\{1,\nu\}$.

We will denote by boldface letters quaternions whose components are vectors (lowercase letters) or matrices (capital letters) of the same dimensions.

2.2. Second-Order Statistics and Properness of Quaternion Random Vectors

Very often, it is necessary to study the characteristics of a quaternion signal by analyzing the second-order statistics. However, second-order information within a quaternion vector $\mathbf{q}$ cannot be estimated only from its correlation matrix $\mathbf{C_{qq}} = \mathrm{E}\{\mathbf{qq}^H\}$, but we need to define the complementary covariance matrices that augment the information within the covariance: $\mathbf{C_{qq}}_{\hat{\imath}} = \mathrm{E}\{\mathbf{qq}^{\hat{\imath}H}\}$,

Table 1. Properties of $\mathbb{Q}$-proper random variables.

$\mathrm{E}\{\mathbf{q}_\delta^2\} = \mathrm{E}\{\mathbf{q}_\epsilon^2\} = \sigma^2$	$\forall \delta, \epsilon = \{a,b,c,d\}$
$\mathrm{E}\{\mathbf{q}_\delta \mathbf{q}_\epsilon\} = \mathbf{0}$	$\forall \delta, \epsilon = \{a,b,c,d\}, \delta \neq \epsilon$
$\mathrm{E}\{\mathbf{qq}\} = -2\sigma^2$	$\forall \delta = \{a,b,c,d\}$
$\mathrm{E}\{|\mathbf{q}|^2\} = 4\sigma^2$	$\forall \delta = \{a,b,c,d\}$

$\mathbf{C_{qq}}_{\hat{\jmath}} = \mathrm{E}\{\mathbf{qq}^{\hat{\jmath}H}\}$, $\mathbf{C_{qq}}_{\hat{\kappa}} = \mathrm{E}\{\mathbf{qq}^{\hat{\kappa}H}\}$ (see also [20], among others, for further details). Thus, we can introduce the augmented covariance matrix of an augmented quaternion vector $\tilde{\mathbf{q}} = \begin{bmatrix} \mathbf{q}^T & \mathbf{q}^{\hat{\imath}T} & \mathbf{q}^{\hat{\jmath}T} & \mathbf{q}^{\hat{\kappa}T} \end{bmatrix}^T$, as:

$$
\tilde{\mathbf{C}}_{\mathbf{qq}} = \mathrm{E}\{\tilde{\mathbf{q}}\tilde{\mathbf{q}}^H\} = \begin{bmatrix}
\mathbf{C_{qq}} & \mathbf{C_{qq}}_{\hat{\imath}} & \mathbf{C_{qq}}_{\hat{\jmath}} & \mathbf{C_{qq}}_{\hat{\kappa}} \\
\mathbf{C_{qq}}_{\hat{\imath}}^H & \mathbf{C}_{\mathbf{q}^{\hat{\imath}}\mathbf{q}^{\hat{\imath}}} & \mathbf{C}_{\mathbf{q}^{\hat{\imath}}\mathbf{q}^{\hat{\jmath}}} & \mathbf{C}_{\mathbf{q}^{\hat{\imath}}\mathbf{q}^{\hat{\kappa}}} \\
\mathbf{C_{qq}}_{\hat{\jmath}}^H & \mathbf{C}_{\mathbf{q}^{\hat{\jmath}}\mathbf{q}^{\hat{\imath}}} & \mathbf{C}_{\mathbf{q}^{\hat{\jmath}}\mathbf{q}^{\hat{\jmath}}} & \mathbf{C}_{\mathbf{q}^{\hat{\jmath}}\mathbf{q}^{\hat{\kappa}}} \\
\mathbf{C_{qq}}_{\hat{\kappa}}^H & \mathbf{C}_{\mathbf{q}^{\hat{\kappa}}\mathbf{q}^{\hat{\imath}}} & \mathbf{C}_{\mathbf{q}^{\hat{\kappa}}\mathbf{q}^{\hat{\jmath}}} & \mathbf{C}_{\mathbf{q}^{\hat{\kappa}}\mathbf{q}^{\hat{\kappa}}}
\end{bmatrix}
\qquad (5)
$$

It is now possible to introduce the quaternion-valued second-order circularity, or $\mathbb{Q}$-properness.

Definition 1 ($\mathbb{Q}$-Properness) *A quaternion random vector* $\mathbf{q}$ *is $\mathbb{Q}$-proper iff the three complementary covariance matrices* $\mathbf{C_{qq}}_{\hat{\imath}}$, $\mathbf{C_{qq}}_{\hat{\jmath}}$ *and* $\mathbf{C_{qq}}_{\hat{\kappa}}$ *vanish:*

$$\mathrm{E}\{\mathbf{qq}^{\hat{\imath}H}\} = \mathbf{0} \quad \mathrm{E}\{\mathbf{qq}^{\hat{\jmath}H}\} = \mathbf{0} \quad \mathrm{E}\{\mathbf{qq}^{\hat{\kappa}H}\} = \mathbf{0}. \quad (6)$$

The properness implies that $\mathbf{q}$ is not correlated with its vector involutions $\mathbf{q}^{\hat{\imath}}, \mathbf{q}^{\hat{\jmath}}, \mathbf{q}^{\hat{\kappa}}$. The main properties of a $\mathbb{Q}$-proper random variable are collected in Table 1 [20], where σ^2 denotes the variance of $\mathbf{q}$.

A quaternion-valued random variable is Gaussian if all its components are jointly normal (see [20], among others), i.e., $p(\tilde{\mathbf{q}}) = p(\mathbf{q}, \mathbf{q}^{\hat{\imath}}, \mathbf{q}^{\hat{\jmath}}, \mathbf{q}^{ik})$, thus the Gaussian probability density function for an augmented multivariate quaternion-valued random vector $\tilde{\mathbf{q}}$ is expressed as:

$$p(\tilde{\mathbf{q}}) = \frac{\exp\left\{-\frac{1}{2}(\tilde{\mathbf{q}} - \tilde{\boldsymbol{\mu}})^H \tilde{\mathbf{C}}_{\mathbf{qq}}^{-1}(\tilde{\mathbf{q}} - \tilde{\boldsymbol{\mu}})\right\}}{(\pi/2)^{2N} \det\left(\tilde{\mathbf{C}}_{\mathbf{qq}}\right)^{1/2}}, \qquad (7)$$

where $\tilde{\boldsymbol{\mu}}$ is the augmented quaternion-valued mean vector for $\tilde{\mathbf{q}}$. For a $\mathbb{Q}$-proper random vector, using the properties of Table 1 and replacing (6) in (5), the augmented covariance matrix $\tilde{\mathbf{C}}_{\mathbf{qq}}$ becomes equal to $4\sigma^2\mathbf{I}$ (see also [20]), and the mean value refers directly to the quaternion vector, i.e., $\boldsymbol{\mu}$, thus we obtain a simplified expression of the Gaussian distribution:

$$p(\tilde{\mathbf{q}}) = \frac{1}{(2\pi\sigma^2)^{2N}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{q} - \boldsymbol{\mu})^H(\mathbf{q} - \boldsymbol{\mu})\right\} \quad (8)$$

where the argument of the exponential is a real function of only $|\mathbf{q} - \boldsymbol{\mu}|^2 = (\mathbf{q} - \boldsymbol{\mu})^{\mathrm{H}} (\mathbf{q} - \boldsymbol{\mu})$.

3. VARIATIONAL AUTOENCODER IN THE QUATERNION DOMAIN

Here, we introduce the novel quaternion variational autoencoder (QVAE) as a generative method based on the probabilistic relation between the quaternion-valued input space and the quaternion-valued latent space. The QVAE aims at controlling the distribution of the quaternion-valued latent vector, which has the characteristic of being a $\mathbb{Q}$-proper quaternion-valued random vector.

3.1. Variational Inference in the Quaternion Domain

Let us consider a quaternion-valued latent vector $\mathbf{z} \in \mathbb{H}$, characterized by a prior probability distribution $p_{\boldsymbol{\theta}}(\mathbf{z})$, and a quaternion input $\mathbf{x} \in \mathbb{H}$, whose conditional probability density function with respect to $\mathbf{z}$ is expressed as $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$. Similarly to the real-valued VAE [1], the QVAE introduces an approximation $\hat{p}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ of the true posterior distribution, thus it is possible to express the marginal likelihood as:

$$
\begin{aligned}
\log\left(p_{\boldsymbol{\theta}}(\mathbf{x})\right) &= \lambda \mathcal{D}_{\mathrm{KL}}\left(\hat{p}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \,||\, p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})\right) + \mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}\right) \\
&= -\lambda \mathcal{D}_{\mathrm{KL}}\left(\hat{p}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \,||\, p_{\boldsymbol{\theta}}(\mathbf{z})\right) + \mathrm{E}_{\hat{p}}\left\{\log\left(p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})\right)\right\}
\end{aligned} \quad (9)
$$

where $\mathcal{D}_{\mathrm{KL}}(\cdot)$ denotes the Kullback-Leibler (KL) divergence and $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x})$ is the variational lower bound with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ [1, 25]. The parameter λ is used to better scale KL values with respect to $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x})$.

We assume that both the recognition model $\hat{p}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ and the generative model $p_{\boldsymbol{\theta}}(\mathbf{z})$ are based on $\mathbb{Q}$-proper Gaussian distributions, i.e., they follow the rule in (8). In particular, the distribution of $\hat{p}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ can be characterized by a quaternion-valued mean vector $\boldsymbol{\mu}_{\mathbf{z}}$ and by an augmented covariance matrix $\tilde{\mathbf{C}}_{\mathbf{zz}}$. This matrix is a block-diagonal matrix, whose submatrices on the main diagonal, extending the properties of the $\mathbb{Q}$-proper random variables in Table 1, are equal and show the same variance, thus $\tilde{\mathbf{C}}_{\mathbf{zz}} = \mathrm{diag}\{\boldsymbol{\Sigma}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}\}$, where $\boldsymbol{\Sigma} = \mathrm{diag}\{\boldsymbol{\sigma}_{\mathbf{z}}^2\}$ and $\boldsymbol{\sigma}_{\mathbf{z}}^2$ is the quaternion-valued vector containing the variance values for each quaternion component. The quaternion-valued mean and variance vectors, $\boldsymbol{\mu}_{\mathbf{z}}$ and $\boldsymbol{\sigma}_{\mathbf{z}}^2$ respectively, are computed by using each a quaternion-valued single-layer neural network.

On the other hand, the prior $p_{\boldsymbol{\theta}}(\mathbf{z})$ is assumed to be a centered isotropic $\mathbb{Q}$-proper Gaussian distribution, i.e., $p_{\boldsymbol{\theta}}(\mathbf{z}) \backsim \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I})$. These considerations allow to approximate the expectation in (9) as $\mathrm{E}_{\hat{p}}\left\{\log\left(p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})\right)\right\} \approx \frac{1}{L}\sum_{l=1}^{L} \log\left(p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}_l)\right)$, where $\mathbf{z}_l$, with $l = 1, \dots, L$, denotes the first L samples drawn from $\hat{p}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$.

Once created the $\mathbb{Q}$-proper Gaussian prior distribution, we perform a reparametrization trick similarly to the real-valued VAE [1], thus having $\mathbf{z} = \boldsymbol{\mu}_{\mathbf{z}} + \boldsymbol{\sigma}_z \odot \boldsymbol{\epsilon}$, where $\odot$ denotes an element-wise product and $\boldsymbol{\epsilon} \in \mathbb{H} \backsim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Considering that both $\hat{p}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ and $p_{\boldsymbol{\theta}}(\mathbf{z})$ are $\mathbb{Q}$-proper distributions and that $\tilde{\mathbf{C}}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} = 4\mathbf{I}$, $\boldsymbol{\mu}_{\boldsymbol{\epsilon}} = \mathbf{0}$ being $p_{\boldsymbol{\theta}}(\mathbf{z}) \backsim \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I})$, the quaternion-valued KL loss in (9) can be expressed as [19, 26]:

$$
\begin{aligned}
\mathcal{D}_{\mathrm{KL}}\left(\hat{p}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \,||\, p_{\boldsymbol{\theta}}(\mathbf{z})\right) &= \frac{1}{2}\left(\mathrm{Tr}\left\{\tilde{\mathbf{C}}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}^{-\frac{1}{2}}\tilde{\mathbf{C}}_{\mathbf{zz}}\tilde{\mathbf{C}}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}^{-\frac{1}{2}}\right\}\right. \\
&\left. + (\boldsymbol{\mu}_{\boldsymbol{\epsilon}} - \boldsymbol{\mu}_{\mathbf{z}})^{\mathrm{H}} \tilde{\mathbf{C}}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\epsilon}} - \boldsymbol{\mu}_{\mathbf{z}}) - N + \log\left(\frac{\det\left(\tilde{\mathbf{C}}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}\right)}{\det\left(\tilde{\mathbf{C}}_{\mathbf{zz}}\right)}\right)\right) \\
&= \frac{1}{2}\left(\mathrm{Tr}\{\boldsymbol{\Sigma}_{\mathbf{z}}\} + \boldsymbol{\mu}_{\mathbf{z}}^{\mathrm{H}}\boldsymbol{\mu}_{\mathbf{z}} - N\right) - 2\sum_{i=1}^{N}\log\left(\sigma_i^2\right). \quad (10)
\end{aligned}
$$

It is worth noting that the minimization of the KL divergence (10) provides a measure of $\mathbb{Q}$-improperness [19].

3.2. Network Architecture

For the scope of the paper, here we use a rather simple architecture in order to prove the benefits of the variational inference in the quaternion domain. Thus, we consider an encoder network composed of quaternion convolutional layers.

The quaternion convolution is one of the main operations of deep neural networks in the quaternion domain [10, 12]. Considering a generic quaternion input vector, $\mathbf{q}$, defined similarly to (1), and a generic quaternion filter matrix defined as $\mathbf{W} = \mathbf{W}_a + \mathbf{W}_b\hat{\imath} + \mathbf{W}_c\hat{\jmath} + \mathbf{W}_d\hat{\kappa}$, the quaternion convolution can be expressed as the following Hamilton product:

$$
\begin{aligned}
\mathbf{W} \otimes \mathbf{q} &= (\mathbf{W}_a\mathbf{q}_a - \mathbf{W}_b\mathbf{q}_b - \mathbf{W}_c\mathbf{q}_c - \mathbf{W}_d\mathbf{q}_d) \\
&\quad + (\mathbf{W}_a\mathbf{q}_b + \mathbf{W}_b\mathbf{q}_a + \mathbf{W}_c\mathbf{q}_d - \mathbf{W}_d\mathbf{q}_c)\hat{\imath} \\
&\quad + (\mathbf{W}_a\mathbf{q}_c - \mathbf{W}_b\mathbf{q}_d + \mathbf{W}_c\mathbf{q}_a + \mathbf{W}_d\mathbf{q}_b)\hat{\jmath} \\
&\quad + (\mathbf{W}_a\mathbf{q}_d + \mathbf{W}_b\mathbf{q}_c - \mathbf{W}_c\mathbf{q}_b + \mathbf{W}_d\mathbf{q}_a)\hat{\kappa}
\end{aligned} \quad (11)
$$

The Hamilton product of (11) allows quaternion neural networks to capture internal latent relations within the features of a quaternion. Each quaternion convolutional layer is followed by a split quaternion Leaky-ReLU activation function [12]. Quaternion batch normalization is not taken into account as it may be considered as a source of randomness that could cause instability [5,6]. Then, as also said in the previous subsection, a quaternion-valued fully-connected output layer is added to the encoder to achieve the quaternion-valued augmented covariance matrix that is used to sample the quaternion latent variable $\mathbf{z}$ and compute the KL divergence loss.

Similarly to the encoder, for the decoder network we use quaternion transposed convolutional layers, each one followed by a quaternion Leaky-ReLU activation function.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed method on the CelebFaces Attributes Dataset (CelebA) [27], a large-scale

Table 2. Averaged results from objective metrics on reconstruction (SSIM, MSE) and generation (FID) tasks. Scores can be read as follows: the higher the better for SSIM, the lower the better for MSE and FID.

	SSIM	MSE	FID	# parameters
VAE	0.8492	0.0047	195.7	3,762,539
QVAE	**0.8941**	**0.0031**	**175.7**	**1,404,996**

face attribute dataset with $202,599$ images that we crop and scale to 64×64 pixels, as in [28]. We investigate the behavior of the proposed QVAE in comparison with a plain VAE in both reconstruction and generation tasks. We report generated samples from both the models for a visual evaluation, as well as results in terms of structural similarity index measure (SSIM), mean-square error (MSE) and Fréchet Inception Distance (FID), for a more objective appraisal.

We consider an encoder with 5 convolutional blocks for the VAE and quaternion-convolutional layer for the proposed QVAE, both with dimensions $(32, 64, 128, 256, 512)$, and a transposed convolutional decoder $(512, 256, 128, 64, 32)$ (equivalently, quaternion transposed convolutions in the proposed method). Both the networks consider Adam optimizer with an initial learning rate of 0.0005 decreased by a factor of 0.5, similarly as [28]. The dimensionality of the latent space is set to 100 in the quaternion domain[1]. We consider a loss composed of binary cross entropy (BCE) for reconstruction and a KL divergence weighted by $\lambda = 0.00001$.

Due to the quaternion operations, the QVAE network has a significantly lower number of parameters with respect to the real-valued VAE (about less than half of the parameters, as shown in Table 2), thus gaining considerable memory advantages. However, despite the lighter architecture in terms of parameters, the proposed QVAE clearly outperforms the real-valued method in terms of objective metrics for the reconstruction task. As shown in Table 2, the QVAE scores a significantly better value both for SSIM and for MSE.

Figure 4 reports the reconstructed images compared with the original ones. On one hand, VAE generates samples focusing just on the face and leaving hair, neck and background very blurred, almost indistinguishable. On the other hand, the proposed QVAE is able to generate samples considering both face and background. Indeed, even hair and background are reconstructed more similarly to the ground truth samples, thus producing more realistic images overall.

Concerning the generation, which is usually the most challenging task, we report both the generated sample sets and the results in terms of objective metric. Figure 4 shows the generated images from the plain VAE and the ones sampled from the proposed QVAE network. While the first set

Fig. 1. Original test set and reconstructed samples sets from plain VAE and proposed QVAE.

Fig. 2. Generated fake image samples from the plain VAE and the proposed QVAE.

seems to have a more various background, the generated faces are less detailed and sometimes confused with the environment. On the contrary, the set sampled from the QVAE shows less heterogeneous background, but significantly more accurate face contours and, overall, a stabler generation. The superior generation ability of the proposed QVAE is underlined also by the results in Table 2 in terms of the FID score. The FID computes the distance of the statistics of generated samples with real ones, thus lower FID values correspond to more real generated samples.

Both reconstruction and generation results prove the effectiveness of defining a latent space and operating in the quaternion domain, while having a significant reduction of the overall number of network parameters.

5. CONCLUSION

In this paper, we proposed a novel approach for variational autoencoders, characterized by the definition of the generative model in the quaternion domain. In particular, the proposed QVAE is able to learn latent representations in the quaternion domain by leveraging the augmented second-order statistics of the quaternion-valued input. Moreover, the QVAE involves the quaternion convolutional layers in both the encoder and the decoder networks, which lead to a significant reduction of the overall number of network parameters. We considered a plain QVAE to clearly show the benefits of the new quaternion-based approach. Results have shown the effectiveness of the proposed approach in both reconstruction and generation tasks with respect to the real-valued counterpart. Future works will extend the proposed QVAE approach to more complex and advanced deep generative models.

6. REFERENCES

[1] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv Preprint: arXiv:1312.6114v10*, pp. 1–14, May 2014.

[2] D. J. Rezende, L. Metz, and S. Chintala, "Stochastic backprop-agation in approximate inference in deep generative models," in *Int. Conf. on Machine Learning (ICML)*, Beijing, China, June 2014, pp. 1278–1286.

[3] A. Razavi, A. van den Oord, B. Poole, and O. Vinyals, "Pre-venting posterior collapse with δ-VAE," in *Int. Conf. on Learning Representations (ICLR)*, New Orleans, LA, May 2019, pp. 1–24.

[4] A. Vahdat, E. Andriyash, and W. G. Macready, "Undirected graphical models as approximate posteriors," in *Int. Conf. on Machine Learning (ICML)*, Vienna, Austria, July 2020, pp. 2266–2275.

[5] L. Maaløe, M. Fraccaro, V. Liévin, and O. Winder, "BIVA: A very deep hierarchy of latent variables for generative mod-eling," in *Advances in Neural Information Process. Systems (NIPS)*, Vancouver, Canada, Dec. 2019, pp. 6548–6558.

[6] A. Vahdat and J. Kautz, "NVAE: A deep hierarchical varia-tional autoencoder," *arXiv Preprint: arXiv:2007.03898v1*, pp. 1–20, July 2020.

[7] T. Bülow and G. Sommer, "Hypercomplex signal – A novel extension of the analystic signal to the multidimensional case," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2844–2852, Nov. 2001.

[8] D. P. Mandic, C. Jahanchahi, and C. Cheong Took, "A quater-nion gradient operator and its applications," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 47–50, Jan. 2011.

[9] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "Frequency-domain adaptive filtering: From real to hypercom-plex signal processing," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 7745–7749.

[10] C. Gaudet and A. Maida, "Deep quaternion networks," in *IEEE Int. Joint Conf. on Neural Netw. (IJCNN)*, Rio de Janeiro, Brazil, July 2018.

[11] T. Parcollet, M. Ravanelli, M. Morchid, G. Linarès, C. Tra-belsi, R. De Mori, and Y. Bengio, "Quaternion recurrent neu-ral networks," in *Int. Conf. Learning Representations (ICLR)*, New Orleans, LA, May 2019, pp. 1–19.

[12] T. Parcollet, M. Morchid, and G. Linarès, "A survey of quater-nion neural networks," *Artif. Intell. Rev.*, Aug. 2019.

[13] R. Vecchi, S. Scardapane, D. Comminiello, and A. Uncini, "Compressing deep-quaternion neural networks with targeted regularisation," *CAAI Trans. Intell. Technol.*, vol. 5, no. 3, pp. 172–176, Sept. 2020.

[14] T. Parcollet, M. Morchid, X. Bost, G. Linarès, and R. De Mori, "Real to H-space autoencoders for theme identification in tele-phone conversations," *IEEE/ACM Trans. Audio, Speech, Lan-guage Process.*, vol. 28, pp. 198–210, 2020.

[15] D. Comminiello, M. Lella, S. Scardapane, and A. Uncini, "Quaternion convolutional neural networks for detection and localization of 3D sound events," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 8533–8537.

[16] M. Ricciardi Celsi, S. Scardapane, and D. Comminiello, "Quaternion neural networks for 3D sound source localiza-tion in reverberant environments," in *IEEE Int. Workshop on Machine Learning for Signal Process.*, Espoo, Finland, Sept. 2020, pp. 1–6.

[17] T. Parcollet, M. Morchid, and G. Linarès, "Quaternion con-volutional neural networks for heterogeneous image process-ing," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 8514–8518.

[18] T. Parcollet, M. Morchid, G. Linarès, and R. De Mori, "Bidi-rectional quaternion long short-term memory recurrent neu-ral networks for speech recognition," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 8519–8523.

[19] J. Vìa, D. Ramìrez, and I. Santamarìa, "Proper and widely linear processing of quaternion random vectors," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3502–3515, July 2010.

[20] C. Cheong Took and D. P. Mandic, "Augmented second-order statistics of quaternion random signals," *Signal Process.*, vol. 91, no. 2, pp. 214–224, Feb. 2011.

[21] J. Vìa, L. Palomar, D. P. Vielva, and I. Santamarìa, "Quater-nion ICA from second-order statistics," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1586–1600, Apr. 2011.

[22] F. Ortolani, D. Comminiello, and A. Uncini, "The widely lin-ear block quaternion least mean square algorithm for fast com-putation in 3D audio systems," in *26th IEEE Workshop on Ma-chine Learning for Signal Process. (MLSP)*, Vietri sul Mare, Italy, Sept. 2016.

[23] N. Le Bihan, "The geometry of proper quaternion random vari-ables," *Signal Process.*, vol. 138, pp. 106–116, Sept. 2017.

[24] H. Sadeghi, E. Andriyash, W. Vinci, L. Buffoni, and M. H. Amin, "PixelVAE++: Improved PixelVAE with discrete prior," *arXiv Preprint: arXiv:1908.09948v1*, pp. 1–12, Aug. 2019.

[25] D. M. Blei, M. I. Jordan, and J. W. Paisley, "Variational Bayesian inference with stochastic search," in *Int. Conf. on Machine Learning (ICML)*, Edinburgh, UK, June 2012, pp. 1367–1374.

[26] C. Doersch, "Tutorial on variational autoencoders," *arXiv Preprint: arXiv:1606.05908v2*, pp. 1–23, Aug. 2016.

[27] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *IEEE Int. Conf. on Computer Vision (ICCV)*, Santiago, Chile, Dec. 2015, pp. 3730–3738.

[28] X. Hou, L. Shen, K. Sun, and G. Qiu, "Deep feature consistent variational autoencoder," in *IEEE Winter Conf. on Applications of Computer Vision (WACV)*, Santa Rosa, CA, Mar. 2017, pp. 1133–1141.