

A connection between image processing and artificial neural networks layers through a geometric model of visual perception

Thomas Batard¹, Eduard Ramon Maldonado^{1,2}, Gabriele Steidl³, and
Marcelo Bertalmío⁴

¹ Crisalix, Barcelona, Spain {thomas.batard,eduard.ramon@crisalix.com}

² Universitat Politècnica de Catalunya, Barcelona, Spain

³ TU Kaiserslautern, Kaiserslautern, Germany {steidl@mathematik.uni-kl.de}

⁴ Universitat Pompeu Fabra, Barcelona, Spain {marcelo.bertalmio@upf.edu}

Abstract. In this paper, we establish a connection between image processing, visual perception, and deep learning by introducing a mathematical model inspired by visual perception from which neural network layers and image processing models for color correction can be derived. Our model is inspired by the geometry of visual perception and couples a geometric model for the organization of some neurons in the visual cortex with a geometric model of color perception. More precisely, the model is a combination of a Wilson-Cowan equation describing the activity of neurons responding to edges and textures in the area V1 of the visual cortex and a Retinex model of color vision. For some particular activation functions, this yields a color correction model which processes simultaneously edges/textures, encoded into a Riemannian metric, and the color contrast, encoded into a nonlocal covariant derivative. Then, we show that the proposed model can be assimilated to a residual layer provided that the activation function is nonlinear and to a convolutional layer for a linear activation function. Finally, we show the accuracy of the model for deep learning by testing it on the MNIST dataset for digit classification.

Keywords: Differential Geometry · Variational Model · Image processing · Vision · Neural Network.

1 Introduction

Deep learning techniques based on artificial neural networks (ANNs) provide state-of-the-arts results in many computer vision tasks [18]. The original goal of the ANN approach was to solve problems in the same way as a human brain. However, over time, attention moved to performing specific tasks, leading to deviations from biology. Nonetheless, some layers like the convolutional layers [25] and the residual units [19],[20], which are at the core of some efficient ANNs, combine linear and nonlinear operators mimicking the neuronal activity (see e.g. [27] and references therein).

The linear operator involved in ANNs is mainly the standard (Euclidean) convolution operator. One of the success of the convolution operator for deep learning tasks is, besides its simplicity which enables fast computation, its ability to generate different types of filters, like isotropic/anisotropic and low-pass/high-pass. Another key property satisfied by the convolution operator is the fact that it commutes with translations, as ANNs satisfying invariance/equivariance with respect to group transformations are desirable in many computer vision tasks like the ones involving recognition or classification. Starting with the seminal convolutional neural network (CNN) [25] involving the standard convolution operator, ANNs enlarging the symmetry group to the Euclidean transformations have been developed by applying the convolution on the special Euclidean Lie group $SE(2)$ [2] or some of its subgroups [12, 13].

Based on the following two properties of the convolution on $SE(2)$, we are led to propose a new operator in the context of ANNs. First, the convolution on $SE(2)$ is related to the interactions between neurons located at different orientation columns in the area V1 of the visual cortex. However, this convolution does not take into account the specificity of these interactions, with the presence of horizontal connections between these neurons and a function modeling neurons activation. Then, the convolution on $SE(2)$ requires the lifting of the visual input from the 2D domain to $SE(2)$. Nonetheless, no lifting is required to reach Euclidean invariance, as this latter can be obtained by means of well-chosen differential operators on the 2D domain. Hence, our proposal is to first, consider a model describing the activity of a population of neurons responding to spatial features in V1 called the Wilson-Cowan equations [29, 30, 7, 10, 14], then to derive an operator from a 2D version of the Wilson-Cowan equations.

The main contribution of this paper is two-folds. On one hand, we establish a connection between the proposed operator and ANNs by showing that a linear activation of the neurons yields a convolutional layer corresponding to a polynomial of order 1 of a differential operator as in [22, 23], whereas a nonlinear activation yields a residual unit [19]. On the other hand, we establish a connection with image processing by showing that, for some nonlinear activation function, the proposed operator is related to a variational model refining the color correction models in [4, 1], and a linear activation yields a quadratic relaxation of the variational model.

In Section 2, we first present a recent form of the Wilson-Cowan equations for neurons responding to edges and textures and its main properties [10], [14]. Then, we introduce a 2D simplification of this model by making use of a cell selectivity principle [11]. Finally, we show for some particular activation functions, that the proposed 2D simplification corresponds to the gradient descent equations of variational models for image processing refining the one in [4]. In Section 3, we improve the model developed in Section 2 by coupling the simplified Wilson-Cowan equations with a model of color perception introduced in [1] and inspired by the Retinex theory of color vision [24]. For particular activation functions, we show that the coupling model represents the gradient descent equations of variational models refining the one in [1].

In Section 4, we first derive an operator from the variational models developed in Section 3 and establish a connection between this operator and ANNs. In particular, we show that a linear activation function yields an accurate convolution operator by showing the ability of this operator to generate different types of local filters and commute with isometries. Finally, we apply this operator in the context of deep learning by inserting it into a simple ANN and test this latter on the MNIST dataset for digit classification.

2 From Wilson-Cowan equations to image processing

2.1 Wilson-Cowan equations encoding the edges and textures

It has been shown by Hubel and Wiesel that there exist neurons responding selectively to the local orientation of the visual input in the area V1 of the visual cortex [21]. The Wilson-Cowan equations [29, 30, 7, 8] describe the temporal evolution of the mean activity of these neurons. More recently, the existence of populations of neurons in V1 encoding the structure tensor $\mathcal{T} \in \text{SPD}(2, \mathbb{R})$, see [15], and organized as a column at each cortical position has been suggested by Chossat and Faugeras [10]. Here, $\text{SPD}(2, \mathbb{R})$ denotes the cone of symmetric positive definite matrices equipped with the affine invariant metric and corresponding distance function $d_{\text{SPD}(2, \mathbb{R})}$. Based on this and inspired by the Wilson-Cowan equations, an evolution equation describing the temporal activity of such neurons has been proposed in [8]. Treating V1 as a planar sheet Ω , the evolution of the mean activity $a: \Omega \times \text{SPD}(2, \mathbb{R}) \times [0, \infty) \rightarrow \mathbb{R}$ of a population of cells with cortical coordinates $x \in \mathbb{R}^2$ and structure tensor preference \mathcal{T} can be modeled (for low firing rates) with the following integro-differential equation (*generalized Wilson-Cowan equations*)

$$\begin{aligned} \frac{\partial a(x, \mathcal{T}, t)}{\partial t} = & -\alpha a(x, \mathcal{T}, t) + h(x, \mathcal{T}, t) \\ & + \int_{\Omega} \int_{\text{SPD}(2, \mathbb{R})} w((x, \mathcal{T}), (y, \mathcal{T}')) \sigma(a(y, \mathcal{T}', t)) d\mathcal{T}' dy, \end{aligned} \quad (1)$$

where $\alpha \geq 0$, σ is an activation function, h the visual input and w the synaptic weights.

2.2 Properties of the equations.

On the synaptic weights. Following the decomposition of the synaptic weights in the original Wilson-Cowan equations into vertical and horizontal terms [9], Faye and Chossat [14] suggested that the synaptic weights in (1) are of the form

$$w((x, \mathcal{T}), (y, \mathcal{T}')) = w_{ver}(d_{\text{SPD}(2, \mathbb{R})}(\mathcal{T}, \mathcal{T}')) \delta(x-y) + \lambda w_{hor}^{\kappa}((x, \mathcal{T}), (y, \mathcal{T}')) (1 - \delta(x-y)), \quad (2)$$

which are split into vertical connections w_{ver} between neurons in the same column and horizontal connections w_{hor}^{κ} between neurons in different columns.

Here, λ is a trade-off parameter whose sign indicates whether the horizontal connections have a net excitatory or inhibitory effect. In particular, they propose the horizontal connections of the form

$$w_{hor}^\kappa((x, \mathcal{T}), (y, \mathcal{T}')) = K_\rho \left(\sqrt{(x-y)^T (\mathbb{I}_2 + \kappa \mathcal{T})(x-y)} \right) \times \mathcal{K}(d_{SPD(2, \mathbb{R})}(\mathcal{T}, \mathcal{T}')), \quad (3)$$

where K_ρ is a Gaussian kernel of variance ρ , $\kappa \geq 0$, and \mathcal{K} is an even positive function of compact support.

Equivariance with respect to group transformations. One of the key property of the original Wilson-Cowan equations is their equivariance with respect to $E(2)$ transformations (assuming that h is 0) [7]. Faye and Chossat [14] show that, under some action of $E(2)$ on $\mathbb{R}^2 \times SPD(2, \mathbb{R})$, the Wilson-Cowan equations (1) satisfy an $E(2)$ -equivariance as well. Moreover, in the limit case where $\kappa = 0$, they point out that the equation admits a $GL(2, \mathbb{R})$ symmetry.

On the activation function. As mentioned by Wilson and Cowan [30], the qualitative properties of solutions of the original Wilson-Cowan equations do not depend on the particular form of the activation function σ and it is likely that σ will differ across the neocortex. Then, assuming that the same property holds for the Wilson-Cowan equations (1), and following the way Bertalmío and Cowan rewrote the activation function of the original Wilson-Cowan equation in [6], we replace the term $\sigma[a(y, \mathcal{T}', t)]$ by $\sigma[f(a(x, \mathcal{T}, t), a(y, \mathcal{T}', t))]$ for some real-valued function f .

The activation function σ is often assumed to be nonlinear and having a specific form, i.e. it is taken as a non-decreasing function such that $\sigma(0) = 0$, saturating at the infinity, and shaped as a sigmoid. In this context, a linear function can be seen as the limit of such a function when increasing its stiffness, and linear activations have actually been considered as well (see e.g. [28]).

2.3 Connection to image processing models

Simplifying the Wilson-Cowan equations through cell selectivity. In [11], the visual cortex is abstracted as $\mathbb{R}^2 \times S^1$, so that the columns only encode the orientations. Here, it is supposed that the simple cells sensitive to the orientation of the gradient of the visual input are maximally activated and selected. Assuming that the same property holds for the columns encoding edges and textures, the cell sensitives to the structure tensor \mathcal{T}_h of the visual input h are maximally activated and selected at each cortical position.

Together with the rewriting of the activation function aforementioned, it yields the following *simplified Wilson-Cowan equation*

$$\frac{\partial a(x, t)}{\partial t} = -\alpha a(x, t) + h(x, t) + \lambda \int_{\Omega} w(x, y) \sigma[f(a(x, t), a(y, t))] dy, \quad (4)$$



Fig. 1. Contrast modification model (9). From left to right: Result of the model for $\lambda < 0$ - Original image - Result of the model for $\lambda > 0$.

where the weight function w is of the form

$$w(x, y) = K_\rho \left(\sqrt{(x - y)^T (\mathbb{I}_2 + \kappa \mathcal{T}_h(x)) (x - y)} \right) \times \mathcal{K} \left(d_{SPD(2, \mathbb{R})}(\mathcal{T}_h(x), \mathcal{T}_h(y)) \right). \quad (5)$$

This simplified Wilson-Cowan equation differs from the one in [6]. Indeed, whereas the weight function (5) encodes the structure tensor of the visual input at each cortical position, the simplification in [6] produces weights of the form

$$w(x, y) = K_\rho(\|x - y\|). \quad (6)$$

where $\|\cdot\|$ stands for the Euclidean vector norm.

Image restoration model with a nonlinear activation function. Assuming in the following that the weight w is symmetric, the input h is constant in time, $\alpha = 1$, and the activation function is of the form

$$\sigma[f(a(x, t), a(y, t))] = s_\epsilon(w(x, y)[a(x, t) - a(y, t)]),$$

where s_ϵ is a differentiable approximation of the sign function, the evolution equation (4) can be seen as the gradient descent equation of a differentiable approximation of the variational model

$$\arg \min_a \frac{1}{2} \int_{\Omega} (a(x) - h(x))^2 dx - \frac{\lambda}{2} \int_{\Omega^2} w(x, y) |a(y) - a(x)| dx dy. \quad (7)$$

Note that the weight (5) can be symmetrized in a straightforward way if we replace $\mathcal{T}_h(x)$ by $(\mathcal{T}_h(x) + \mathcal{T}_h(y))/2$ in the first term.

For w of the form (6), we recover the variational model in [4], which realizes contrast enhancement for $\lambda > 0$ and contrast reduction for $\lambda < 0$ provided that the variance ρ is large.

Image restoration model with a linear activation function. Assuming that the activation function is of the form

$$\sigma[f(a(x, t), a(y, t))] = w(x, y)[a(x, t) - a(y, t)],$$

the evolution equation (4) corresponds to the gradient descent of the following variational problem

$$\arg \min_a \frac{1}{2} \int_{\Omega} (a(x) - h(x))^2 dx - \frac{\lambda}{2} \int_{\Omega^2} w^2(x, y) (a(y) - a(x))^2 dx dy. \quad (8)$$

In particular, for w being a normalized Gaussian kernel, the solution of the corresponding Euler-Lagrange equations is

$$a = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(h)}{(1 - \lambda) - \lambda \mathcal{F}(w^2)} \right) \quad (9)$$

where $\mathcal{F}, \mathcal{F}^{-1}$ are the Fourier transform and its inverse, respectively.

Fig. 1 shows the results of applying formula (9) to each channel of the input image h in Fig. 1 (center), where the variance of the Gaussian kernel w is taken large, and for two opposite signs of λ . As in the nonlinear case (7), the model performs contrast enhancement for $\lambda > 0$ (Fig. 1 (right)) and contrast reduction for $\lambda < 0$ (Fig. 1 (left)).

3 Wilson-Cowan equations and color perception models

3.1 A geometric model of color perception

Retinex models of color vision aim to reproduce the perception of the colors of a scene inspired by psychophysical/physiological knowledge about color vision [5, 31]. They can be interpreted as the averaging of perceptual distances between image pixels, as pointed out in [1].

Given an RGB color image $u = (u^1, u^2, u^3)$, the perceived image $L = (L^1, L^2, L^3)$ is, according to Kernel-Based Retinex [5], given for $k = 1, 2, 3$, by

$$L^k(x) = \int_{y: u^k(y) \geq u^k(x)} w(x, y) \left[A \log \left(\frac{u^k(x)}{u^k(y)} \right) + 1 \right] dy + \int_{y: u^k(y) < u^k(x)} w(x, y) dy,$$

where w is a Gaussian kernel and A is a constant, which can be rewritten as

$$\begin{aligned} L^k(x) &= \int_{y \in \Omega} w(x, y) \zeta (\log[u^k(x)] - \log[u^k(y)]) dy \\ &= \int_{y \in \Omega} w(x, y) \zeta \left(\int_{\gamma_{y,x}} \nabla \log(u^k)(\gamma_{y,x}(t)) dt \right) dy \end{aligned} \quad (10)$$

for some nonlinear function ζ , and for any path $\gamma_{y,x}$ joining y to x , see [1]. The quantity $\nabla \log(u^k)$ can be interpreted as the perceived gradient of the image according to Weber's law in vision. However, Weber's law suffers from several limitations, and formula (10) can be improved replacing $\nabla \log(u^k)$ by a more accurate representation of the perceptual gradient.

Based on the assumption that the color constancy property comes from an equivariance of the perceived gradient with respect to light changes, Georgiev [16] suggested that a well-chosen covariant derivative ∇^E on a vector bundle E is a good candidate to describe the perceived gradient, due to the invariance of this differential operator with respect to moving frame changes.

Coupling formula (10) with Georgiev's approach yields the following expression of the perceived image

$$\begin{aligned} L(x) &= \int_{y \in \Omega} w(x, y) \zeta \left(\int_{\gamma_{y,x}} \nabla_{\gamma'_{y,x}(t)}^E u(\gamma_{y,x}(t)) dt \right) dy \\ &= \int_{y \in \Omega} w(x, y) \zeta \left(u(x) - \tau_{y,x,\gamma_{y,x}}^E u(y) \right) dy, \end{aligned} \quad (11)$$

where $\tau_{y,x,\gamma_{y,x}}^E$ denotes the parallel transport map with respect to ∇^E along a path $\gamma_{y,x}$ joining y to x .

Finally, let us mention that the vector bundle framework is consistent with neurophysiological studies suggesting the existence of neurons in V1 responding to colors and organized as a column at each cortical coordinates [21, 26], and with the presence of horizontal connections between color columns suggested in [3]. Indeed, the columns can be assimilated to the fibers of the bundle and the horizontal connections between the columns to the parallel transport map.

3.2 Connection to image processing models

Section 3.1 revealed a close relation between the formula (11) describing the perceived colors of a scene and the integral term in (4). In what follows, we assume that the vector bundle is equipped with a metric $\|\cdot\|$ and a covariant derivative ∇^E compatible with this metric. Again, we assume that $\alpha = 1$, h constant in time and w symmetric.

Image restoration model with a nonlinear activation function. Assuming that the activation function in (4) is of the form

$$\sigma[f(a(x, t), a(y, t))] = s_\epsilon(w(x, y)[a(x, t) - \tau_{y,x,\gamma_{y,x}}^E a(y, t)]),$$

where $s_\epsilon = z/\sqrt{\|z\|^2 + \epsilon}$ is the differentiable approximation of $s(z) = z/\|z\|$, then the evolution equation (4) corresponds to the gradient descent equation of a differentiable approximation of the variational problem

$$\arg \min_a \frac{1}{2} \int_{\Omega} \|a(x) - h(x)\|^2 dx - \frac{\lambda}{2} \int_{\Omega^2} w(x, y) \|\tau_{y,x,\gamma_{y,x}}^E a(y) - a(x)\| dx dy. \quad (12)$$

Our variational model (12) is a refinement of the one for color image correction introduced in [1] as the weight function in this latter is of the form (6), whereas the proposed weight function, which is a symmetrization of the weight (5), encodes the structure tensor, such that our model can take the edges and textures of the image into account.



Fig. 2. Comparison of two contrast enhancement models. From left to right: Original image "Giocondarioca" - Result of the model (9) - Result of the model (14).

Image restoration model with a linear activation function. Assuming that the activation function in (4) is of the form

$$\sigma[f(a(x, t), a(y, t))] = w(x, y)[a(x, t) - \tau_{y, x, \gamma_{y, x}}^E a(y, t)],$$

the evolution equation (4) corresponds to the gradient descent equation of the variational model

$$\arg \min_a \frac{1}{2} \int_{\Omega} \|a(x) - h(x)\|^2 dx - \frac{\lambda}{2} \int_{\Omega^2} w^2(x, y) \|\tau_{y, x, \gamma_{y, x}}^E a(y) - a(x)\|^2 dx dy. \quad (13)$$

In the particular case where the covariant derivative is flat and the weight w is a normalized Gaussian kernel, the solution of the Euler-Lagrange equations is

$$a = P \left(\mathcal{F}^{-1} \left(\frac{\mathcal{F}(P^{-1}h)}{(1 - \lambda) - \lambda \mathcal{F}(w^2)} \right) \right), \quad (14)$$

where P is the moving frame in which the covariant derivative is trivial.

Fig. 2 compares the model (14) for a flat covariant derivative encoding some brightness perception phenomenon, called the Helmholtz-Kohlrausch effect (see Sect. 5.5.2 in [1] for the expression of the moving frame P), to its Euclidean restriction (9). Both models are applied with the same parameter $\lambda > 0$ and the same large variance for the Gaussian kernel. We observe that the model (14) provides a better result as it preserves more the colors of the original image.

4 Connection to artificial neural networks

4.1 A new operator and its connections to existing layers

The proposed operator. An ANN can be described as a sequence of functions $\{a^l\}$, $l = 0, \dots, L$, with $a^l: \Omega^l \rightarrow \mathbb{R}^{n_l}$ and operators $\{H^l\}$, such that $a^{l+1} =$

$H^l(a^l)$. Hence, an operator in an ANN should be able to modify the dimension of its input. Then, together with the coupling between the simplified Wilson-Cowan equation (4) and the color perception model (11), we derive the following layer

$$H^l(a^l): x \longmapsto W^l a^l(x) + Q^l \int_{\Omega^l} w(x, y) \beta(w(x, y) [\tau_{y,x,\gamma_{y,x}}^{E^l} W^l a^l(y) - W^l a^l(x)]) dy, \quad (15)$$

where β is antisymmetric, $W^l \in \text{End}(\mathbb{R}^{n_l}; \mathbb{R}^{n_{l+1}})$, $Q^l \in \text{End}(\mathbb{R}^{n_{l+1}})$ and τ^{E^l} is the parallel transport map associated to a covariant derivative ∇^{E^l} .

Connection to convolutional layers. It is common to call convolutional any layer where the operator H^l is linear. Then, assuming that $\beta = Id$, the operator (15) writes

$$H^l(a^l) = W^l a^l + Q^l \Delta_w^{E^l NL} W^l a^l, \quad (16)$$

where the operator $\Delta_w^{E^l NL} := \frac{1}{2} \nabla_w^{E^l NL*} \nabla_w^{E^l NL}$ corresponds to an extension of the (Euclidean) nonlocal Laplacian [17] to vector bundles, called generalized nonlocal Laplacian, with

$$\nabla_w^{E^l NL} u: (x, y) \longmapsto w(x, y) [\tau_{y,x,\gamma_{y,x}}^E u(y) - u(x)]$$

being the nonlocal covariant derivative induced by a covariant derivative ∇^E and a weight function w , and

$$\nabla_w^{E^l NL*} \eta: x \longmapsto \int_{\Omega} w(x, y) [\tau_{y,x,\gamma_{y,x}}^E \eta(y, x) - \eta(x, y)] dy$$

its adjoint (see [1] for more details).

The layer (16) is a polynomial of degree 1 of a differential operator, as in the graph convolutional networks [22] and surface networks [23] approaches.

Connection to residual layers. A residual layer is of the form

$$a^{l+1} = W^l a^l + F^l(a^l), \quad (17)$$

where the so-called residual function F^l combines linear and nonlinear operators.

Then, we can observe that, for β nonlinear, the operator (15) can be considered as a residual layer where

$$F^l = Q^l \nabla_w^{E^l NL*} \beta \nabla_w^{E^l NL} W^l.$$

4.2 Learning the parameters of the operator.

In deep learning, the parameters of the key layers of an ANN (e.g. convolutional or residual layers) are learned on a well-chosen training data, this latter depending on the task to be performed (classification, recognition, denoising, etc). In what follows, we show that the parameters of the operator (15) have to satisfy some constraints in order for this operator to be consistent in the context of deep learning.

Avoid Overfitting. One of the main constraint when designing an ANN for deep learning is to avoid overfitting, as this latter makes the ANN memorize rather than learn the training data, and which greatly affects the results on the test data afterwards. Overfitting can occur if the number of trainable parameters is too big. In the CNN approach [25], this problem is addressed by convolving the signal with a spatially invariant template of small support. Then, in the context of the operator (15), this problem can be addressed in a two-folds way: first, by reducing the spatial support of the weight function; then, by making the structure tensor \mathcal{T}^l and the connection 1-forms ω^{E^l} be constant on the whole domains Ω^l . This makes the weight function be layer dependent and of the form

$$w^l(x, y) = K_\rho \left(\sqrt{(x - y)^T (\mathbb{I}_2 + \kappa \mathcal{T}^l) (x - y)} \right)$$

for some structure tensor \mathcal{T}^l , where the inner term corresponds to the distance between x and y induced by the metric $g^l := \mathbb{I}_2 + \kappa \mathcal{T}^l$.

Generate different types of local filters. It is crucial to make the key layers of an ANN be able to generate a large variety of filters. In the CNN approach, this is achieved by learning the template to which the signal is convolved. Moreover, the spatial invariance of the template makes the convolutional layer be equivariant with respect to translations, and consequently makes the neural network satisfy some invariance with respect to translations, which is a desirable property in many computer vision tasks.

We claim that the operator (15) can generate a great variety of filters as well by making the parameters $\mathcal{T}^l, \omega^{E^l}, W_l, Q_l$ be trainable. For instance, the operator can be a low-pass or a high-pass filter, depending on the values of W_l and Q_l . Moreover, the metric g^l determines whether the filter is isotropic ($g^l \equiv \mathbb{I}_2$) or anisotropic ($g^l \neq \mathbb{I}_2$), and invariance with respect to the isometries of (Ω^l, g^l) can be achieved for (15) of the form (16) and $\Delta_w^{E^l NL}$ reduced to its local form. Finally, the parallel transport maps τ^{E^l} make the operators H^l mix the channels, yielding new types of filters.

4.3 Experiments

We test the operator (15) for the task of digit classification on the MNIST dataset. The experiment consists in, given an ANN well-designed for this task, replacing the key layer in this ANN by the operator (15), and test the subsequent ANN for three different configurations of the operator.

The ANN we consider here is inspired by the one available in <https://github.com/keras-team/keras>, and stacks the following layers: 1 convolution (32 features), 1 activation (LeakyReLU), 1 maxpooling, 1 convolution (64 features), 1 activation (LeakyReLU), 1 maxpooling, 1 convolution (64 features), 1 activation (LeakyReLU), 1 maxpooling, 1 dropout, 1 flatten, 1 dense, 1 activation (ReLU), 1 dropout, 1 dense. Here, we consider the form (16) of the operator.

Table. 1 shows the mean scores after 50 epochs over 10 tests for each configuration of the operator, where the size of the spatial support of the weight w is 3×3 . The chosen optimizer is Adam with learning rate parameter 10^{-3} . We observe that the best scores are obtained when both the metric and the connection 1-form are learned, which corresponds to the most accurate configuration in terms of visual perception.

Table 1. Mean scores of the proposed operator tested on the MNIST dataset.

Configuration of the operator	Score after 50 epochs	Highest score
Euclidean ($g \equiv Id, \omega^E \equiv 0$)	0.9927	0.9935 (after 39 epochs)
Vector bundle 1 ($g \equiv Id$, learning ω^E)	0.9929	0.9938 (after 38 epochs)
Vector bundle 2 (learning g and ω^E)	0.9933	0.9944 (after 39 epochs)

5 Conclusion

We showed that an operator inspired by vision neuroscience and psychophysics connects image processing models and some layers in artificial neural networks. In order to show the accuracy of this operator for deep learning tasks, we tested it on a simple dataset for classification. Further work will be devoted to insert this operator into deeper ANNs for more complex computer vision and image processing tasks like denoising and deblurring. Moreover, the proposed geometric framework enables to design ANNs on non-Euclidean domains in order to address tasks involving geometric data like 3D shapes and graphs.

References

1. Batard, T., Bertalmío, M.: A Geometric Model of Brightness Perception and its Application to Color Images Correction. *J. Math. Imag. Vis.*, 60(6), pp. 849-881 (2018).
2. Bekkers, E.J., Lafarge, M.W., Veta, M., Eppenhof, K.A.J., Pluim, J.P.W., Duits, R.: Roto-Translation Covariant Convolutional Networks for Medical Image Analysis. *arXiv.1804.03393* (2018).
3. Ben-Shahar and Zucker, S.W.: Hue Geometry and Horizontal Connections. *Neural Networks - Special Issue Vision and Brain*, 17(5-6), pp. 753-771 (2004).
4. Bertalmío, M., Caselles, V., Provenzi, E., Rizzi, A.: Perceptual Color Correction through Variational Techniques. *IEEE Trans. Im. Proc.*, 16(4), pp. 1058-1072 (2007).
5. Bertalmío, M., Caselles, V., Provenzi, E.: Issues about Retinex Theory and Contrast Enhancement. *Int. J. Comput. Vis.*, 83(1), pp. 101-119 (2009).
6. Bertalmío, M. and Cowan J.D.: Implementing the Retinex Algorithm with Wilson-Cowan Equations. *Journal of Physiology-Paris*, 103(1-2), pp. 69-72 (2009).
7. Bressloff, P.C., Cowan, J.D., Golubitsky, M., Thomas, P.J., Wiener, M.C.: Geometric Visual Hallucinations, Euclidean symmetry and the functional architecture of striate Cortex. *Phil. Trans. Roy Soc. Lond. B*, 356, pp. 299-330 (2001).
8. Bressloff, P.C., Cowan, J.D., Golubitsky, M., Thomas, P.J., Wiener, M.C.: What Geometric Visual Hallucinations tell us about the Visual Cortex. *Neural Computation*, 14(3), pp. 473-491 (2002).

9. Bressloff, P.C., Cowan, J.D.: The Functional Geometry and Local and Horizontal Connections in a Model of V1. *Journal of Physiology-Paris*, 97(2-3), pp. 221-236 (2003).
10. Chossat, P., Faugeras, O.: Hyperbolic Planforms in Relation to Visual Edges and Textures Perception. *PLoS Computational Biology*, 5(12), pp. 1-16 (2009).
11. Citti, G., Sarti, A.: A Gauge Field of Modal Completion. *J. Math. Imag. Vis.*, 52(2), pp. 267-284 (2015).
12. Cohen, T.S., Welling, M.: Group Equivariant Convolutional Networks. *Proc. Int. Conf. on Machine Learning ICML* (2016).
13. Dieleman, S., De Fauw, J., Kavukcuoglu, K.: Exploiting Cyclic Symmetry in Convolutional Neural Networks. *Proc. Int. Conf. on Machine Learning ICML* (2016).
14. Faye, G., Chossat, P.: A Spatialized Model of Visual Texture Perception using the Structure Tensor Formalism. *Networks and Heterogeneous Media*, 8(1), pp. 211-260 (2013).
15. Förstner, W., Gülch, E.: A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: *Proc. ISPRS Intercommission Conf. on Fast Processing of Photogrammetric Data*, pp. 281-305 (1987).
16. Georgiev, T.: Relighting, Retinex theory, and Perceived Gradients. *Proceedings of Mirage 2005*.
17. Gilboa, G., Osher, S.: Nonlocal Operators with Applications to Image Processing. *Multiscale Model. Simul.*, 7(3), pp. 1005-1028 (2008).
18. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press (2016).
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognition CVPR* (2016).
20. Huang, G., Zhuang, L., van der Maaten, L., Weinberger, K.Q.: Densely Connected Convolutional Networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognition CVPR* (2017).
21. Hubel, D.H.: *Eye, Brain and Vision*. Scientific American Library, W.H. Freeman & Co (1988).
22. Kipf, T.N., Welling, M.: Semi-supervised Classification with Graph Convolutional Networks. *Proc. Int. Conf. on Learning Representation ICLR* (2017).
23. Kostrikov, I., Jiang, Z., Panozzo, D., Zorin, D., Bruna, J.: Surface Networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognition CVPR* (2017).
24. Land, E., McCann, J.J.: Lightness and Retinex Theory. *J. Optical Soc. of Am.*, 61(1), pp. 1-11 (1971).
25. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based Learning applied to Document Recognition. *Proceedings of the IEEE*, 86(11), pp. 2278-2324 (1998).
26. Livingstone, M.S., and Hubel, D.H.: Anatomy and Physiology of a Color System in the Primate Visual Cortex. *J. Neuroscience*, 4(1), pp. 309-356 (1984).
27. Martinez-Garcia, M., Cyriac, P., Batard, T., Bertalmío, M., Malo, J.: Derivatives and Inverse of Cascaded Linear+Nonlinear Neural Models. *PLOS ONE* 13(10): e0201326 (2018).
28. Neves, L.L., Monteiro, L.H.A: A Linear Analysis of Coupled Wilson-Cowan Neuronal Populations. *Computational Intelligence and Neuroscience* (2016).
29. Wilson, H.R., Cowan, J.D.: Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. *Biophysical Journal*, 12(1), pp. 1-24 (1972).
30. Wilson, H.R., Cowan, J.D.: A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Nervous Tissue. *Kybernetik*, 13(2), pp. 55-80 (1973).
31. Yeonan-Kim, J., Bertalmío, M.: Analysis of Retinal and Cortical Components of Retinex Algorithms. *Journal of Electronic Imaging*, 26(3), 031208 (2017).