Choquet Capacity networks for random point process classification

Mehran Mohammadi, Santiago Velasco-Forero, François Willot, Mateus Sangalli, Jesus Angulo

Choquet Capacity networks for random point process classification and regression

Mehran Mohammadi, Santiago Velasco-Forero*,
François Willot, Mateus Sangalli and Jesus Angulo

Center for Mathematical Morphology, Mines Paris PSL,
35 rue Saint Honoré, 77305 Fontainebleau, France.

Ref: Proceedings in Mathematics and Statistics Vol. 457 (2024) pp. 229–241, Springer. The present document is identical to the published version. Compared to HAL-version 1, a few typos have been corrected and, following the reviewers suggestions, comments have been added (see end of Sec. 3.1). Results unchanged.

Abstract

In this study, we revisit the Choquet capacity in the framework of convolutional neural networks, in (max, +)-algebra. By incorporating a discrete and learnable Choquet capacity model, we enhance the ability to represent the spatial arrangement and density variations in random point processes of convolutional neural networks. To validate the effectiveness of our approach, numerical experiments are conducted on synthetic datasets simulating diverse spatial point patterns of the Neyman-Scott process. When compared to classical convolutional neural networks, the proposed approach exhibits comparable or improved performances in terms of classification. Superior results are also observed in regression problems involving the Neyman-Scott parameter that monitors the point patterns spatial dispersion.

*Corresponding author, santiago.velasco@minesparis.psl.eu

1 Introduction

There is a significant demand for analytical techniques that can handle vast amounts of image data in particular in engineering science [1]. For instance, images from satellites, geological maps, microscope images of metals and materials, cellular tissue, and data from computerized tomography [2] all require automatic and quantitative methods due to the sheer volume of data involved. In some scenarios, the information observed is set of coordinates in 2D or 3D, and the feature of interest depends on the spatial relationship between these points. Examples include individual brain activation maps [3, 4], RNA Localization patterns [5], positions on random sensor networks [6], realizations of random processes [7], among others. On the one hand, mathematical morphology [8, 9, 10] offers ideas that are applicable in this context, particularly due to its interpretation in terms of the size/shape relationship of the object of interest. Cord and co-authors combined textural descriptors and statistical learning to address segmentation problems [11] and extract defects merged in complex real textures [12]. On the other hand, data-driven learning methods have changed the way image problems are solved. Especially the deep models allow to solve recognition problems with performance sometimes superior to humans [13]. In this paper, we follow the results of [14, 15, 16, 17], to present a new application of morphological models learned from data for the characterization of random point processes. The Neyman-Scott point process is presented in section 2. The proposed Choquet Capacity model is described in section 3. Section 4 demonstrate the benefits of the proposed model in comparison with classical convolutional neural networks (CNNs) for classification and regression problems on images generated by Neyman-Scott point process. We conclude in section 5.

2 Neyman-Scott point processes

In the present work, we consider Neyman-Scott point processes [18, 19, 20, 21] as a way to generate point patterns that are statistically different from the classical Poisson model of points [22, 23, 24]. The Neyman-Scott model, like Poisson's one, is a particular type of Boolean random set [25, 26, 8, 27]. Poisson point processes on $\mathbb{R}^2$ depend on a single parameter θ, called the *intensity*. The intensity θ represents the mean number number of points per unit area. In the Poisson point model, the number of points in any two disjoint domains are independent random variables which lead to the following Poisson-distributed law for the number of points $\eta(\Omega)$ in a domain

Ω,

$$\mathbb{P}\{\eta(\Omega) = k\} = \frac{(\theta|\Omega|)^k}{k!} \exp^{-\theta|\Omega|}, \tag{1}$$

where $|\Omega|$ is the area of Ω. This leads to the formula for the *Choquet capacity* $T(K)$ of the Poisson point process $X = \{x_1, x_2, ...\}$:

$$T_X(K) := \mathbb{P}(X \cap K \neq \emptyset) = 1 - \mathbb{P}\{K \subseteq X^c\} = 1 - \exp^{-\theta|K|}, \tag{2}$$

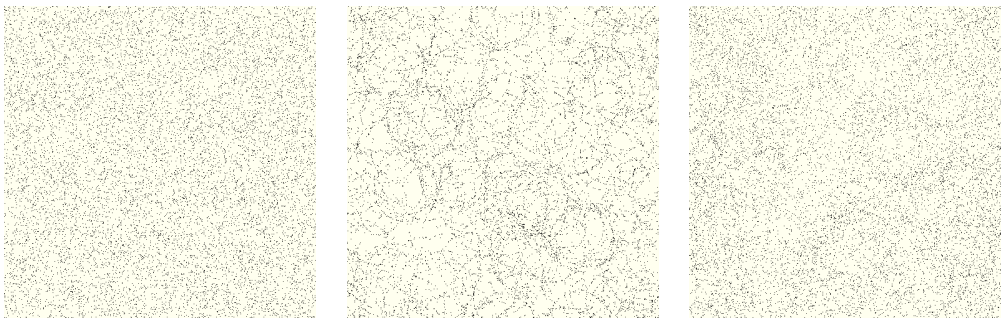
where K is a compact set [28].

Neyman-Scott point processes are different from Poisson point processes in that they generate *clustered* sets of points. This is achieved in two steps. In the first step, a standard Poisson point process of intensity θ_p is generated, leading to a set of *parent points* $x^{(p)} \in \mathbb{R}^2$ ($p \in \mathbb{N}$). In the second step, we associate each parent point with a set of n_p *daughter points* $x_1^{(p)}, \ldots, x_{n_p}^{(p)}$. The Neyman-Scott point process is made of all daughter points; parent points are not included. Neyman-Scott models are accordingly parametrized by θ_p, the probability law for the n_p and that of the daughter points, given in term of the multivariate probability law for the $x_i^{(p)} - x^{(p)}$. For simplicity, we consider a constant distribution $n_p \equiv n$. We also assume that the vectors $x_i^{(p)} - x^{(p)}$ are i.i.d. variables drawn from a multivariate random distribution in $\mathbb{R}^2$, with uniform angular distribution in $[0; 2\pi]$. Distances from parent to daughter points are presently drawn from a Pareto law with parameters α, r_m such as:

$$\mathbb{P}\{|x_i^{(p)} - x^{(p)}| > r\} = \left(\frac{r_m}{r}\right)^{\alpha} H(r - r_m), \qquad \alpha > 0, \quad r_m > 0, \tag{3}$$

where $H(\cdot)$ is the Heavyside function, equal to one on $\mathbb{R}^+$ and zero elsewhere. Here r_m controls the *typical size of the clusters* whereas the parameter α monitors the *spatial dispersion*. Thus, r has infinite variance when $\alpha \leq 2$ and infinite mean when $\alpha \leq 1$. Examples of realizations are shown in Fig. (1). The points spatial dispersion in the Neyman-Scott model is markedly more heterogeneous when α is large (see Fig. (1b). When α approaches one, the Neyman-Scott and Poisson models are difficult to distinguish (see Figs. 1).

Importantly, the Choquet-Matheron-Kendall theorem established independently by [29, 30] states that any stationary and ergodic random set, such as Poisson or Neyman-Scott point processes, is completely determined by the functional $T(K)$. In the next section, we propose a model that uses (2) to find optimal K in a discrete setting (denoted as K), in the sense of learning from labeled data for supervised problems. Note that this learning process makes use of convolutional layer where $\mathbf{K}$ are optimized in the real set. Accordingly, we propose regularization terms to encourage the network to learn binary kernels.



(a) (b) (c)

Figure 1: Realizations of a Poisson point process (a) and Neyman-Scott models (b-c) with the same density of points. b) Parameters $r_m = 0.1$, $\alpha = 100$. c) $r_m = 0.1$, $\alpha = 3$.

3 Choquet Capacity models

3.1 Discrete random set

Usually random sets are defined in the sense of Euclidean geometry, as subsets of $\mathbb{R}^d$. However, in many applications, as in image processing, the domain is discrete, defined in pixels structures in 2D, $\Omega = \mathbb{Z}^2$. The discrete random set (DRS) X on $\mathbb{Z}^2$ is now defined by $\mathsf{X} := \{(i, j) \in \mathbb{Z}^2 : x_{i,j} = 1\}$. The *discrete-capacity functional* of a random set X is defined by [30]:

$$T_{\mathsf{X}}(\mathsf{K}) := \mathbb{P}(\mathsf{X} \cap \mathsf{K} \neq \emptyset) = 1 - \mathbb{P}\{\mathsf{K} \subseteq \mathsf{X}^c\} := 1 - Q_{\mathsf{X}}(\mathsf{K}) \tag{4}$$

for every $\mathsf{K} \in S$ the collection of all bounded subsets of $\mathbb{Z}^2$. From (4), it is easy to see that: a) $0 \leq T_{\mathsf{X}}(\mathsf{K}) \leq 1$ for every $\mathsf{K} \in S$, b) $T_{\mathsf{X}}(\emptyset) = 0$, and c) $T_{\mathsf{X}}(\mathsf{K}_1) \leq T_{\mathsf{X}}(\mathsf{K}_2), \forall \mathsf{K}_1, \mathsf{K}_2 \in S$ such that $\mathsf{K}_1 \subseteq \mathsf{K}_2$. Other important properties can be shown for DRS [31, 32]. According to the Choquet-Kendall-Matheron theorem, any discrete random set X is uniquely characterized by its discrete-capacity functional $T(\mathsf{K})$, defined over all bounded subsets $\mathsf{K} \in S$ of $\mathbb{Z}^2$. Furthermore, we note that:

$$\mathbb{P}(\mathsf{X} \cap \mathsf{K} \neq \emptyset) = \mathbb{P}(x \in \delta_{\check{\mathsf{K}}}(\mathsf{X})) \tag{5}$$

where $\check{\mathsf{K}} = \{-x; x \in \mathsf{K}\}$ denotes the set K mirrored w.r.t. the origin. The capacity in (4) is accordingly obtained from the morphological dilation $\delta_{\check{\mathsf{K}}}(X)$ of X by K [27, 33].

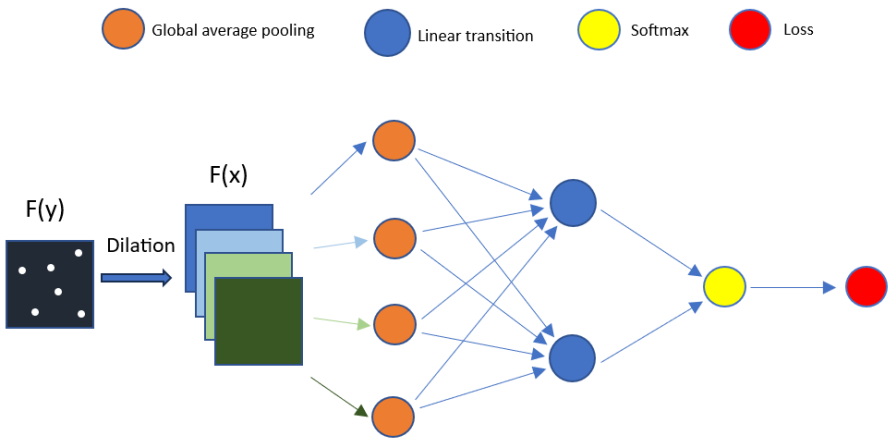


Figure 2: Schematic of the proposed Choquet Capacity model.

Hereafter, the points $(i, j) \in \mathbb{Z}^2$ of the DRS X are obtained from the set X in $\mathbb{R}^2$ so that: $(i, j) \in \mathsf{X}$ whenever $([i; i+1[\times[j; j+1[) \cap X \neq \emptyset$. Accordingly, the DRS X is a partial representation of the Neyman-Scott point process defined by Eq. (3). The representation of X by X is however excellent when $r_m \gg 1$, i.e. when r_m is much larger than the pixel size.

3.2 Choquet Capacity layer

This section describes a layer that can be used for the computation of (5), in the context of neural networks. To do so, we use the extension of convolutional models in the $(\bigvee, +)$ algebra, called max-plus convolutions [34], or *dilation layer* [15]. The dilation layer replaces discrete classical convolutions and can be compared to fuzzy measures in so-called fuzzy integral neural networks [35]. For an input image f, the dilation layer is defined as:

$$\delta_{\mathbf{K}}(f)(x) := \bigvee_{y \in \mathbb{Z}^2} [f(x-y) + \mathbf{K}(y)] = \bigvee_{y \in \mathbb{Z}^2} [f(y) + \mathbf{K}(x-y)] \qquad (6)$$

where $\mathbf{K}$ is a matrix of real values called *kernel* or *structuring function*. In the cases where the image f represents a DRS (denoted by f), and the kernel $\mathbf{K}$ takes on values in $\{-1; 0\}$[1], we have that $\mathsf{f}(y) + \mathbf{K}(x-y) = 0$ if

[1]Hereafter, we consider input images lying in the interval $[0, 1]$, so that in practice, -1 plays the same role as $-\infty$ for kernel.

$\mathbf{K}(x - y) \leq \mathsf{f}(y)$ ($y \in \mathbb{Z}^2$) and 1 elsewhere. By sliding the kernel through the domain and calculating its average, assuming ergodicity, we obtain an estimation for the value of $Q_{\mathsf{f}}(\mathbf{K})$ in (4):

$$\hat{Q}_{\mathsf{f}}(\mathbf{K}) := \frac{1}{M} \sum_{j=1}^{M} \delta_{\mathbf{K}}(\mathsf{f})(x_j) \tag{7}$$

where M is the number of pixels in the image. In the following, equation (7) is called a *Choquet Capacity layer* with convolutional kernels $\mathbf{K}$.

Hereafter, we propose to study the performance of the Choquet Capacity layers to learn n_k kernels, $\mathbf{K}s := \{\mathbf{K}_i\}_{i=1}^{n_k}$, in the framework of neural networks. Accordingly, for each input image f, a vector containing the n_k values computed by (7) is used as features, transformed linearly by a dense layer (or many for multilayer perceptrons (MLP) models) with parameter $\mathbf{W}$, and finally passed trough an activation function adapted to the regression or classification problem. Fig (2) shows a schematic of the proposed Choquet Capacity model for a classification problem. The classical training mechanism on neural networks seeks to find the optimal model's parameters $(\mathbf{W}, \mathbf{K}s)$,

$$\underset{(\mathbf{K}s, \mathbf{W})}{\arg\min} \left\{ \sum_{\ell=1}^{N} \mathtt{loss}(Y_\ell, \widehat{Y_\ell}) + \mathtt{Reg}(\mathbf{W}, \mathbf{K}s) \right\} \tag{8}$$

where $\widehat{Y_\ell}$ is the prediction of the network for the ℓ-th sample, Y_ℓ is the ground truth, N is the number of samples and $\mathtt{Reg}(\mathbf{W}, \mathbf{K}s)$ is a regularization term on the parameters of the model. In its simplest form, the parameters $(\mathbf{W}, \mathbf{K}s)$ are initialized to random values and updated by means of a stochastic gradient descend, the gradient being computed by backpropagation [13]. Note that the shape K in (5) is a bounded subset, whereas each of the matrix $\mathbf{K}$ in (6) is a max-plus convolutional kernel with real values. Therefore, we propose two alternatives that address this difference during the training of the model:

i) *Regularization*: the term $\mathtt{Reg}$ in (8) is set to penalize kernel values that are far from a binary equivalent, in our case of max-plus algebra on binary images, 0 or -1. Two different regularization functions are considered for R:

$$\mathtt{Reg}_1(\mathbf{K}s) = \lambda_1 \sum_{i=1}^{n_k} (\mathbf{K}_i + 1) + \lambda_2 \sum_{i=1}^{n_k} (\mathbf{K}_i + 1)^2, \tag{9a}$$

$$\mathtt{Reg}_2(\mathbf{K}s) = \lambda_1 \sum_{i=1}^{n_k} (\mathbf{K}_i + 1)\mathbf{K}_i + \lambda_2 \sum_{i=1}^{n_k} (\mathbf{K}_i + 1)^2 \mathbf{K}_i^2, \tag{9b}$$

where n_k in the number of kernels and $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are the hyperparameters to control the effect of regularization.

ii) *Quantization* of the kernels during optimization. We use an integrated value filter to enforce kernel values in $[-1, 0]$. We use the *straight-through estimator* proposed in [36, 37]. Predictions are computed by rounding the kernel to the closest integer (-1 or 0) whereas the gradient in the backpropagation is left unchanged.

The different approaches above are compared to one another and to classical convolutional networks in the next section. We would like to study the effect of including more layers into the model. This is called the *depth* of the network in deep learning [38]

4 Numerical experiments

In the present section, we quantify the performances of the various convolutional networks. Two problems are addressed: first, that of classification of realizations of Neyman-Scott point patterns w.r.t. Poisson; second, the prediction of the dispersion parameter α in a regression problem[2].

4.1 Classification of point pattern

In this experiment, we set the image size to 500×500 pixels and $\theta = 0.05$ in number of points per pixel unit. We choose $n = 50$ and let $\theta_p = \theta/n$ so that the density of points is the same irrespective of the model chosen. We also fix $r_m = 0.1 \times 500 = 50$ and let α varies between 1 and 100. All point patterns are statistically homogeneous. We perform the classification task on two classes of Poisson and Neyman-Scott point patterns with different values of the parameter α and conduct a comparative analysis with the CNNs. The data set contains 500 images for each class, with a train-to-validation ratio of 4:1. For simplicity, we use one layer in each neural network, with four filters of size 7×7 pixels. For supervised classification, a common choice of loss function is the *binary cross-entropy* also known as *log loss*, defined by

$$\texttt{logloss}(Y, \widehat{Y}) := -\frac{1}{N} \sum_{\ell=1}^{N} \left[Y_\ell \log q_0^\ell + (1 - Y_\ell) \log q_1^\ell \right] \tag{10}$$

[2]The source code of this section is available at `https://github.com/Jacobiano/ChoquetLayer`

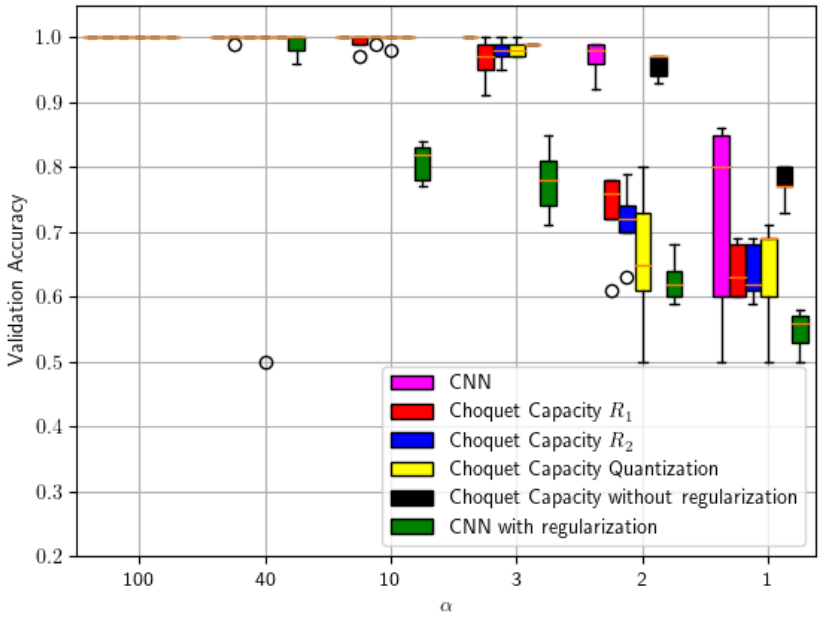
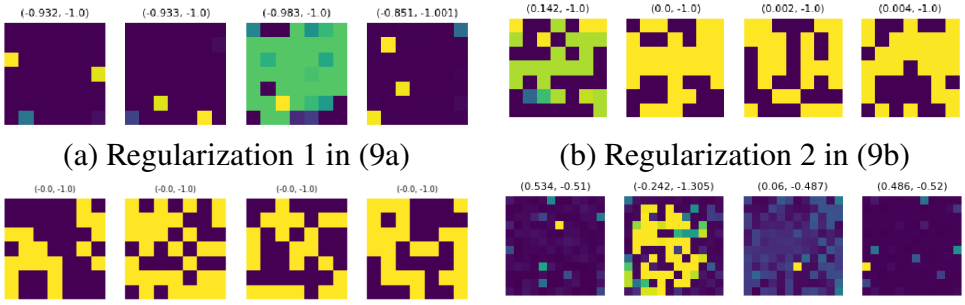


Figure 3: Validation accuracy for the different neural network models in the two-classes classification problem (Poisson and Neyman-Scott point processes) vs. the parameter α that monitors spatial dispersion.

where (q_0^ℓ, q_1^ℓ) denote the prediction of the network for the ℓ-th samples, Y_ℓ the ground truth of the ℓ-th samples and N is the number of samples. The models are trained to predict the class between images produced by Poisson processes and Neyman-Scott models with different values of α parameter. Figure 3 shows the boxplot of validation accuracy for models training from different initialization values. Interestingly, the Choquet Capacity layer demonstrates competitive performance, especially when the parameter α decreases, so that realizations of the Neyman-Scott point pattern are similar to that of a Poisson distribution. When regularization is applied, the model tends to have values $\{-1, 0\}$, but they are still values that prevent us from indicating that the solution is binary. For quantization, $\mathbf{K}s$ contain only values in $\{-1, 0\}$ in Figure (4)(c), although no distinctive shape is observed. However, performance significantly degraded, when regularization or quantization is used, specially in more difficult cases.



(a) Regularization 1 in (9a)

(b) Regularization 2 in (9b)

(c) Quantization operator

(d) No regularization

Figure 4: Example of learned kernel $\mathbf{K}s = \{\mathbf{K}_1, \ldots, \mathbf{K}_4\}$ using different proposed regularization alternatives of the Choquet Capacity layer. Max and min values indicated at the top of each image.

4.2 Parameter regression of the Neyman-Scott point processes

In the present subsection, we compare the performances of different neural network models for predicting the value of the parameter α, in a regression problem. We set the image size to 128×128 pixels, the density of points to $\theta = 100/(128 \times 128)$ and $r_m = 0.02$. We also let α vary between 0 and 9. This last one is the value that we would like to predict. For the regression problem, the *mean absolute error* is the mean absolute difference between true value and prediction:

$$\texttt{MAE}(Y, \widehat{Y}) := -\frac{1}{N} \sum_{\ell=1}^{N} |Y_\ell - \widehat{Y}_\ell| \tag{11}$$

is used as loss function. During training the α values are scaled to the interval $[0, 1]$ as it is standard practice in regression models. As a first example, we compare the result between Choquet Capacity and a classical convolutional layer followed by a multi-layer perceptron that consists of two layers of dimension ten and batch normalization [39], followed by a sigmoid activation, see Figure 5. Note that the convolutional layer and our proposition have exactly the same number of parameters. In this case, our Choquet Capacity model obtains predictions that are clearly quantitatively better.

As a second illustration in the regression case we studied the effect of network depth, prior to the use of the Choquet Capacity model. Usually,

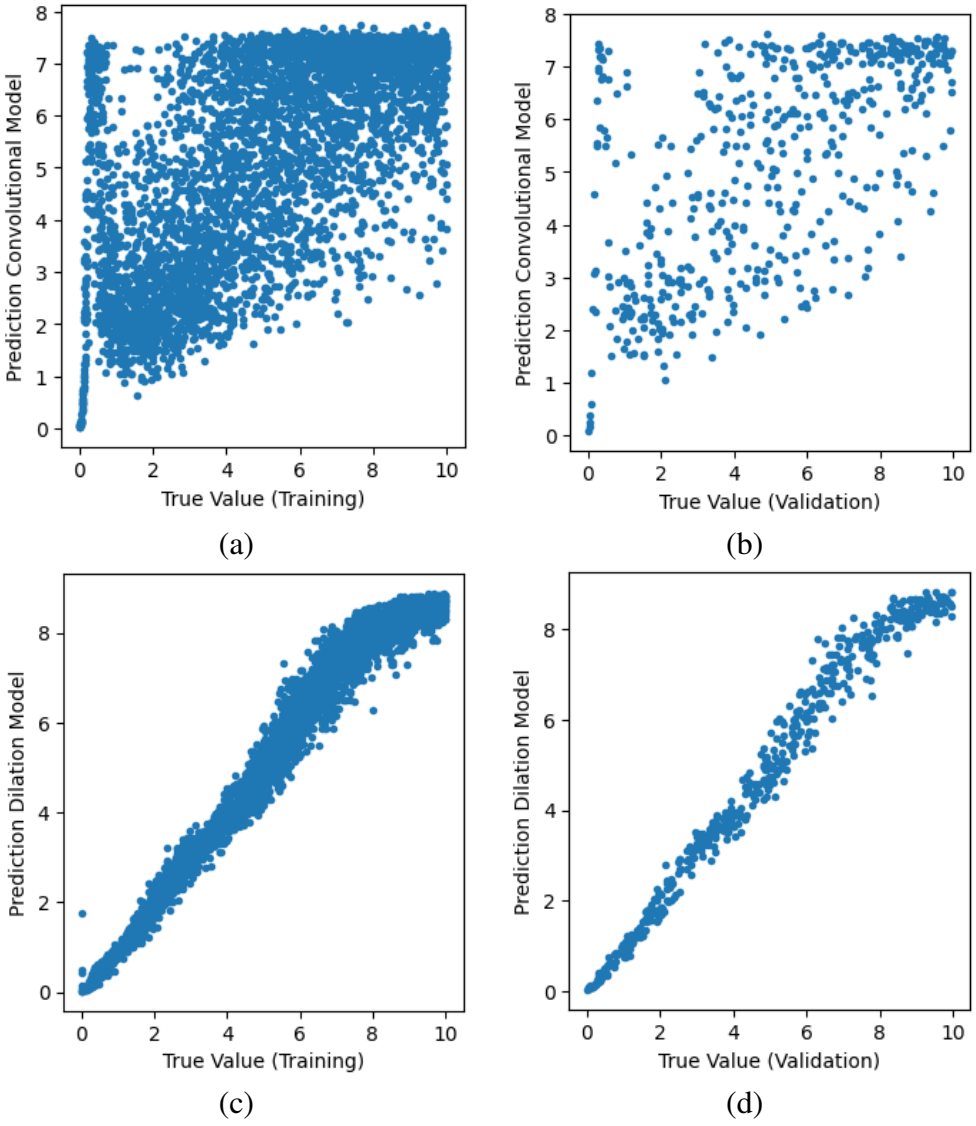


Figure 5: Comparison on the prediction over testing training for one layer convolutional with global average pooling network (a, b) and our Choquet Capacity proposition in the regression problem of the Neyman-Scott point processes (c, d). Models (a) and (c), as well as (b) and (d), have the same number of parameters. All models have been trained with the same protocol.

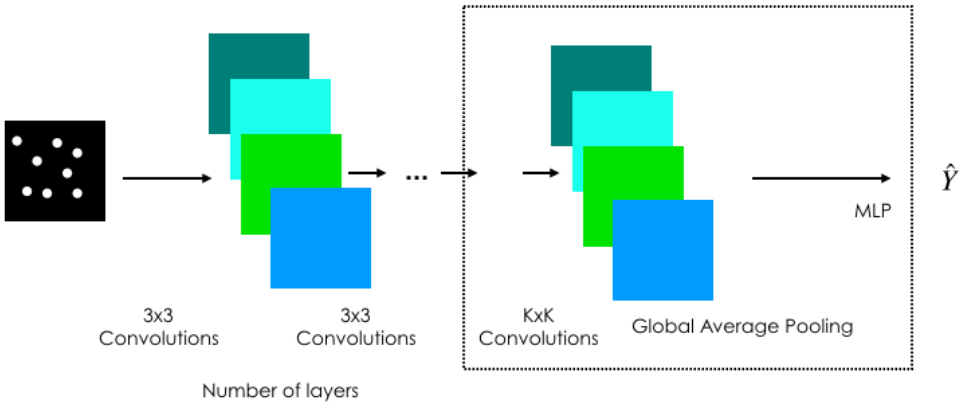


Figure 6: Architecture used in section 4.2, with varying number of layers of classical convolutional layers of kernel size 3×3 pixels followed by ReLU activation.

increasing both depth and width (number of filters per layer) helps to improve the model performance until the number of parameters becomes too high and stronger regularization is needed [40]. For that, we consider convolutional layers with kernel size of 3×3 pixels and 48 filters are concatenated before the use of global pooling and the MLP. This architecture is represented in Figure 6. We compare different convolutions in the last layer before the global pooling: proposed Choquet Capacity model, classical CNN, and depthwise convolutional (DWCNN). The latter is included because it contains exactly the same number of parameters compared to our Choquet Capacity model. In Figure 7, we compare the validation loss of the three networks with the number of parameters in each model shown in logarithmic scale. The number of degrees of freedom is increased by augmenting the depth from one to seven in the convolutional part in each architecture, leading to models with deeper feature extractions. The Choquet Capacity model improves the regression results in all explored cases, for a fixed number of parameters, compared to DWCNN. Finally, Figure 7 also shows than comparable performances to the Choquet model are achieved by classical CNNs at the cost of a much higher number of parameters.

5 Conclusion

The results obtained in this study are based on a novel approach that makes use of dilation layers with global pooling, inspired from the definition of

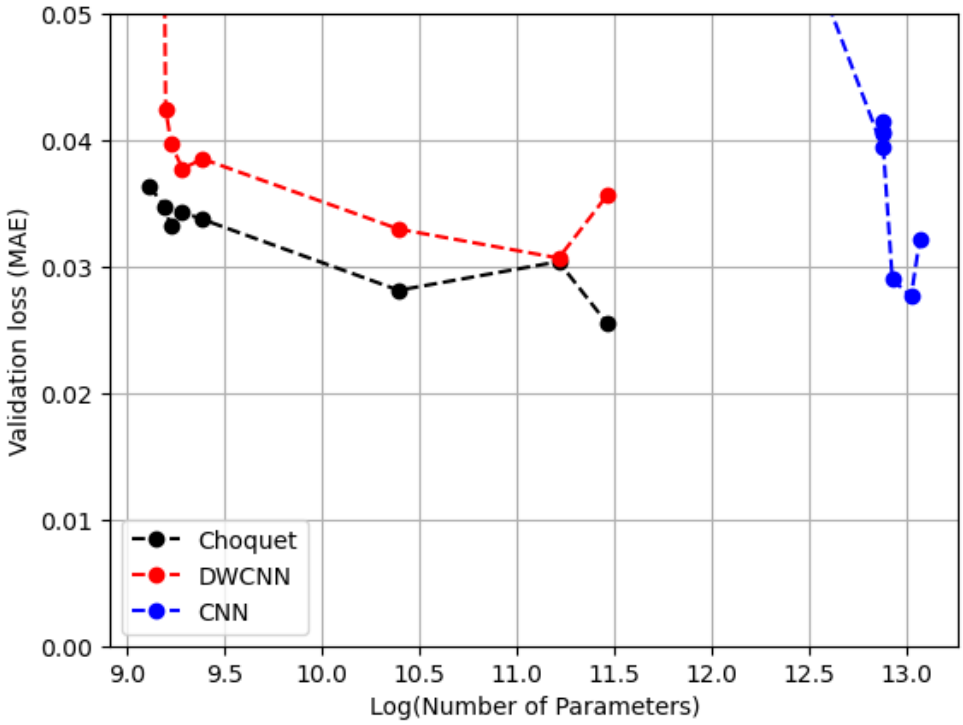


Figure 7: Performance of different neural networks for varying number of layers in each model (See Figure 6). Three architectures are compared. In all cases each layer has 48 kernels of size 3×3, ReLU is used as activation function, and a final layer with a convolution of size 13×13 followed by GlobalAveragePooling, BatchNormalization and a MLP with two layers of dimension ten with a final sigmoid activation. DWCNN refers to a depthwise convolution, CNN to a classical convolution and Choquet Capacity to the max-plus convolutions in 3. DWCNN and Choquet Capacity models have exactly the same number of parameters.

the Choquet capacity of random sets. Numerical experiments performed on synthetic datasets which simulate diverse Neyman-Scott point patterns, demonstrate the efficiency of the method, as well as seemingly superior performances in particular in regression problems involving the dispersion parameter in the Neyman-Scott point process. Compared to classical convolutions, the Choquet Capacity networks are somewhat more difficult to train, due to the presence of a nonlinear convolution filters. Yet they are versatile in that they may be combined with classical convolutional networks. As a negative result, the alternatives studied to motivate the estimation of binary kernels result in very important losses in the performance of the models. This topic, as well as the adequacy of Choquet capacity networks to analyse real-world data, will be addressed in future research works.

Acknowledgements This work was granted access to the HPC resources of IDRIS under the allocation 2023-AD011012212R2 made by GENCI. The authors gratefully acknowledge the financial support of Institut Carnot (grant 220000496) and thank Thomas Walter and Christian Lantuéjoul for helpful discussions.

Competing Interests The authors have no conflicts of interest to declare that are relevant to the content of this chapter.

References

[1] F. Chinesta, E. Cueto, E. Abisset-Chavanne, J. L. Duval, F. E. Khaldi, Virtual, digital and hybrid twins: a new paradigm in data-based engineering and engineered data, Archives of computational methods in engineering 27 (2020) 105–134.

[2] A. Etiemble, T. Douillard, O. Valentin, J.-C. Badot, E. Maire, Numerical prediction of multiscale electronic conductivity of lithium-ion battery positive electrodes, Journal of the Electrochemical Society 166 (8) (2019) A1692.

[3] K. J. Worsley, S. Marrett, P. Neelin, A. Evans, Searching scale space for activation in pet images, Human brain mapping 4 (1) (1996) 74–90.

[4] J.-B. Poline, B. Mazoyer, Analysis of individual brain activation maps using hierarchical description and multiscale detection, IEEE transac. on medical imaging 13 (4) (1994) 702–710.

[5] A. Imbert, F. Mueller, T. Walter, Pointfish: Learning point cloud representations for rna localization patterns, in: ECCV, Springer, 2022, pp. 487–502.

[6] S. S. Ram, D. Manjunath, S. K. Iyer, D. Yogeshwaran, On the path coverage properties of random sensor networks, IEEE Transac. on Mobile Computing 6 (5) (2007) 494–506.

[7] K. Pawlasová, J. Dvořák, Supervised nonparametric classification in the context of replicated point patterns, Image Analysis & Stereology 41 (2) (2022) 57–109.

[8] J. Serra, Image Analysis and Mathematical Morphology, Academic Press Inc, 1983.

[9] P. Soille, Morphological Image Analysis: Principles and Applications, 2nd Edition, Springer-Verlag, Berlin, Heidelberg, 2003.

[10] L. Najman, H. Talbot, Mathematical Morphology: From Theory to Applications, Wiley, 2013.

[11] A. Cord, D. Jeulin, F. R. Bach, Segmentation of random textures by morphological and linear operators, in: Proceedings of the 8$^{\text{th}}$ International Symposium on Mathematical Morphology, Rio de Janeiro, Brazil, Oct. 10–13, 2007, 2007, pp. 387–398.

[12] A. Cord, F. Bach, D. Jeulin, Texture classification by statistical learning from morphological image processing: application to metallic surfaces, Journal of Microscopy 239 (2) (2010) 159–166.

[13] Y. Le Cun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436–444.

[14] M. Sangalli, S. Blusseau, S. Velasco-Forero, J. Angulo, Scale equivariant neural networks with morphological scale-spaces, in: International Conference on Discrete Geometry and Mathematical Morphology, Springer, 2021, pp. 483–495.

[15] S. Velasco-Forero, R. Pagès, J. Angulo, Learnable empirical mode decomposition based on mathematical morphology, SIAM Journal on Imaging Sciences 15 (1) (2022) 23–44.

[16] S. Velasco-Forero, A. Rhim, J. Angulo, Fixed point layers for geodesic morphological operations, in: British Machine Vision Conference, 2022.

[17] R. Mondal, S. Santra, S. S. Mukherjee, B. Chanda, Morphological network: How far can we go with morphological neurons?, in: 33rd

British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022, BMVA Press, 2022.

[18] R. P. Waagepetersen, An estimating function approach to inference for inhomogeneous neyman–scott processes, Biometrics 63 (1) (2007) 252–258.

[19] D. Vere-Jones, Stochastic models for earthquake occurrence, Journal of the Royal Statistical Society: Series B (Methodological) 32 (1) (1970) 1–45.

[20] M. Westcott, Results in the asymptotic and equilibrium theory of poisson cluster processes, Journal of Applied Probability 10 (4) (1973) 807–823.

[21] B. D. Ripley, Modelling spatial patterns, Journal of the Royal Statistical Society: Series B (Methodological) 39 (2) (1977) 172–192.

[22] D. Daley, D. Vere-Jones, An introduction to the theory of point processes: volume I: elementary theory and methods, Second Edition, Springer, 2003.

[23] D. Daley, D. Vere-Jones, An introduction to the theory of point processes: volume II: general theory and structure, Springer Science & Business Media, 2007.

[24] J. Moller, R. P. Waagepetersen, Statistical inference and simulation for spatial point processes, CRC press, 2003.

[25] D. Stoyan, W. S. Kendall, J. Mecke, Stochastic geometry and its applications, John Wiley & Sons, Chichester, 1995, second edition.

[26] D. Jeulin, Morphological models of random structures, Springer, 2021.

[27] J. Serra, Image Analysis and Mathematical Morphology: Vol.: 2: Theoretical Advances, Academic Press, 1988.

[28] G. Choquet, Theory of capacities, in: Annales de l'institut Fourier, Vol. 5, 1954, pp. 131–295.

[29] D. G. Kendall, Foundation of a theory of random sets, Stochastic geometry (1974).

[30] G. Matheron, Random Sets and Integral Geometry, Probability and Statistics Series, Wiley, 1974.

[31] B. Singh, M. Siddiqi, Discrete random functions: Modeling and analysis using mathematical morphology, Mathematical Morphology and its Applications to Image and Signal Processing (1996) 65–72.

[32] J. Goutsias, Morphological analysis of discrete random shapes, Journal of Mathematical Imaging and Vision 2 (1992) 193–215.

[33] D. Jeulin, Random texture models for material structures, Statistics and Computing 10 (2000) 121–132.

[34] M. Akian, R. Bapat, S. Gaubert, et al., Max-plus algebra, Handbook of linear algebra 39 (2006).

[35] M. A. Islam, D. T. Anderson, A. J. Pinar, T. C. Havens, G. Scott, J. M. Keller, Enabling explainable fusion in deep learning with fuzzy integral neural networks, IEEE Transactions on Fuzzy Systems 28 (7) (2019) 1291–1300.

[36] Y. Bengio, N. Léonard, A. Courville, Estimating or propagating gradients through stochastic neurons for conditional computation, arXiv preprint arXiv:1308.3432 (2013).

[37] P. Yin, J. Lyu, S. Zhang, S. Osher, Y. Qi, J. Xin, Understanding straight-through estimator in training activation quantized neural nets, ICLR (2019).

[38] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.

[39] S. Santurkar, D. Tsipras, A. Ilyas, A. Madry, How does batch normalization help optimization?, Advances in neural information processing systems 31 (2018).

[40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE CVPR, 2016, pp. 770–778.