Modification of information reduction processes in Convolutional Neural Networks

PhD dissertation

Iosu Rodriguez-Martinez

PhD Supervisors: Humberto Bustince, Francisco Herrera, Zdenko Takač

Department of Statistics, Computer Science and Mathematics
Public University of Navarre, Spain

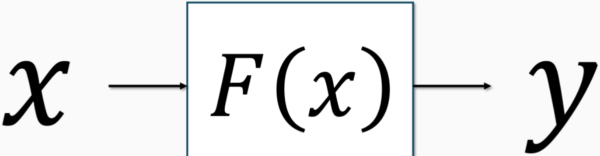
October 11, 2024

Table of contents

upna

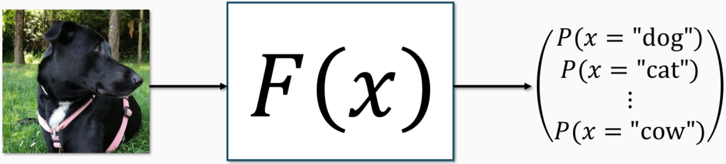
Introduction

What a Neural Network is

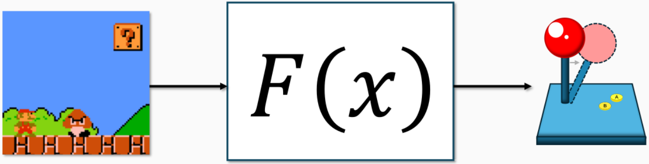


upna

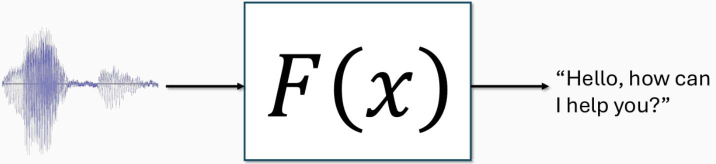
What a Neural Network is



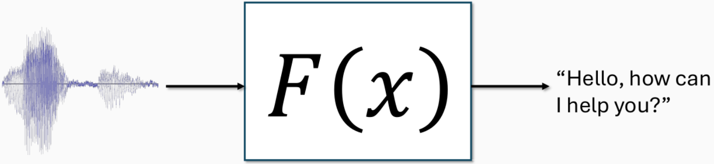
What a Neural Network is



What a Neural Network is

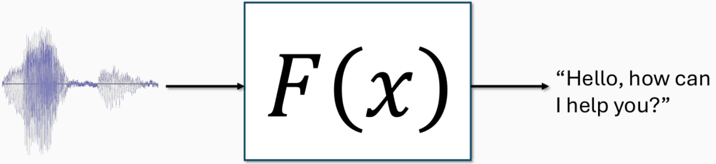


What a Neural Network is



How can we approximate such a function?

What a Neural Network is

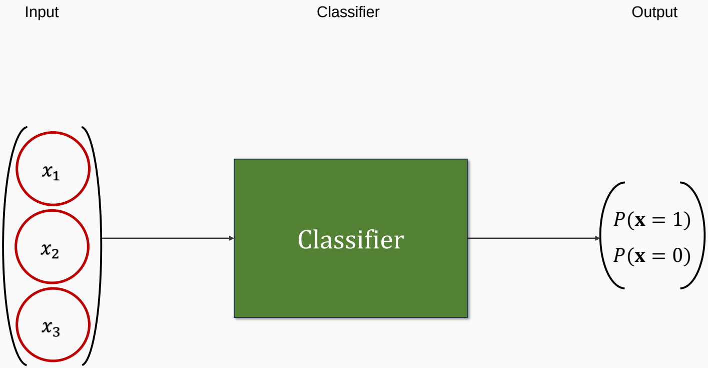


How can we approximate such a function?

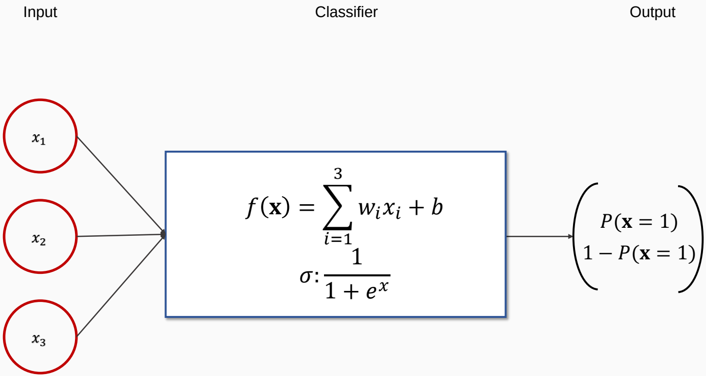
- Set of affine transformations: $\{f^{(l)}, l \in 1, \ldots, L | f^{(l)} : \mathbb{R}^{n_{l-1}} \rightarrow \mathbb{R}^{n_l}\}$
- Non-linear element-wise activation function: $\sigma : \mathbb{R} \rightarrow \mathbb{R}$

$$F(\mathbf{x}) = f^{(L)}\left(\sigma\left(f^{(L-1)}\left(\ldots\left(\sigma\left(f^{(1)}\left(\mathbf{x}\right)\right)\right)\right)\right)\right)$$

The Multilayer Perceptron

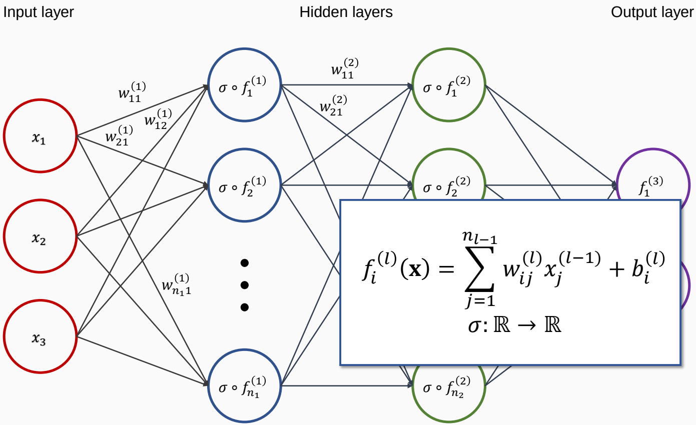


The Multilayer Perceptron



upna

The Multilayer Perceptron

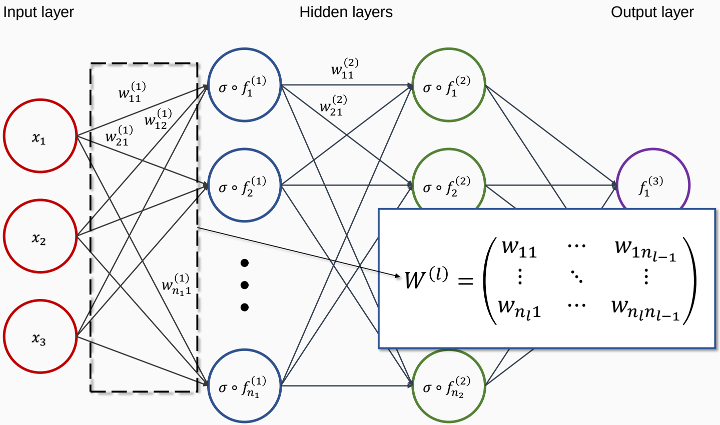


$$f_i^{(l)}(\mathbf{x}) = \sum_{j=1}^{n_{l-1}} w_{ij}^{(l)} x_j^{(l-1)} + b_i^{(l)}$$

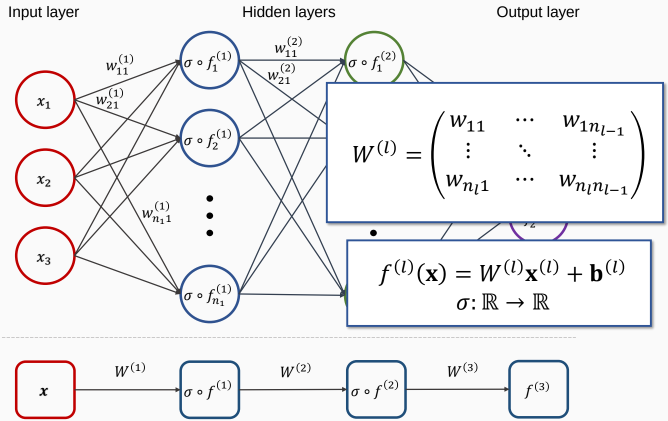
$$\sigma: \mathbb{R} \rightarrow \mathbb{R}$$

Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The expressive power of neural networks: A view from the width. *Advances in neural information processing systems*, 30.

The Multilayer Perceptron



The Multilayer Perceptron

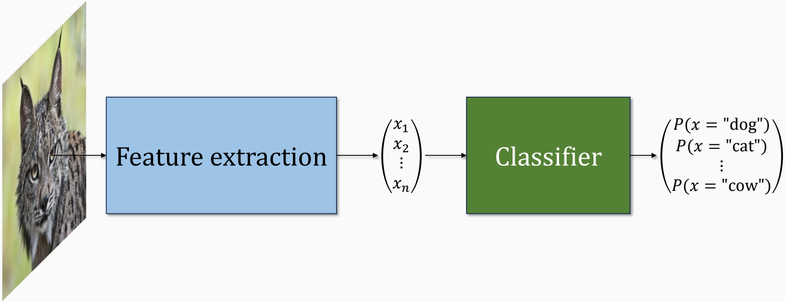


$$W^{(l)} = \begin{pmatrix} w_{11} & \cdots & w_{1n_{l-1}} \\ \vdots & \ddots & \vdots \\ w_{n_l 1} & \cdots & w_{n_l n_{l-1}} \end{pmatrix}$$

$$f^{(l)}(\mathbf{x}) = W^{(l)}\mathbf{x}^{(l)} + \mathbf{b}^{(l)}$$
$$\sigma: \mathbb{R} \to \mathbb{R}$$

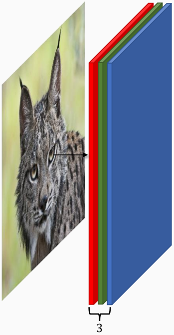
upna

Convolutional Neural Network (CNN)



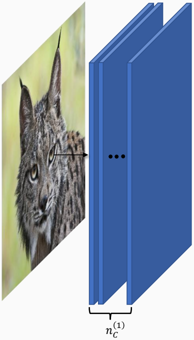
Convolutional Neural Network (CNN)

Feature extraction



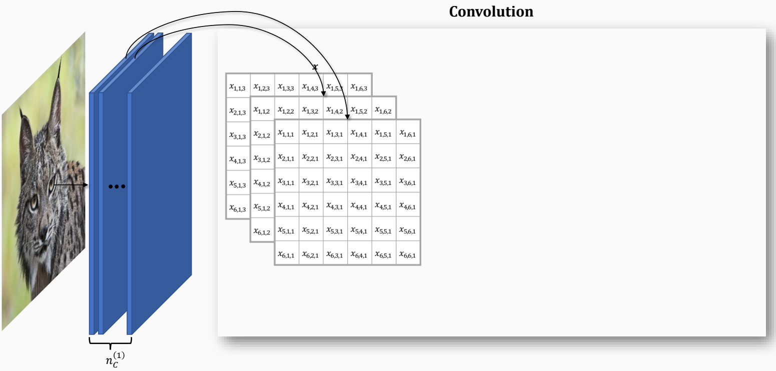
Convolutional Neural Network (CNN)

Feature extraction



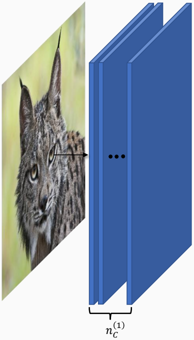
Convolutional Neural Network (CNN)

Feature extraction

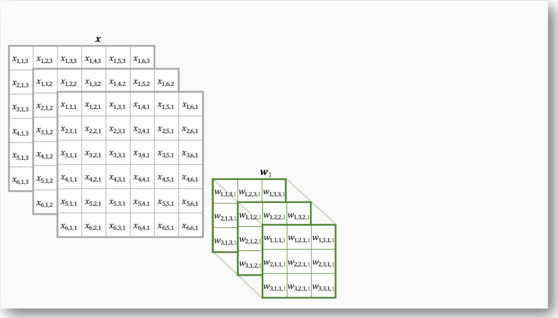


Convolutional Neural Network (CNN)

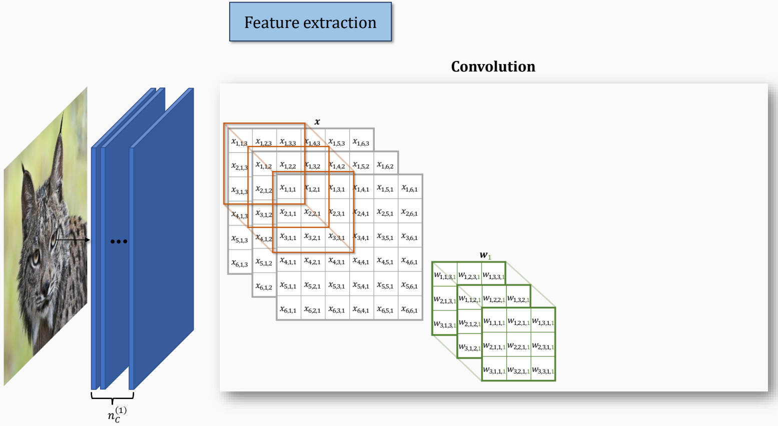
Feature extraction



Convolution



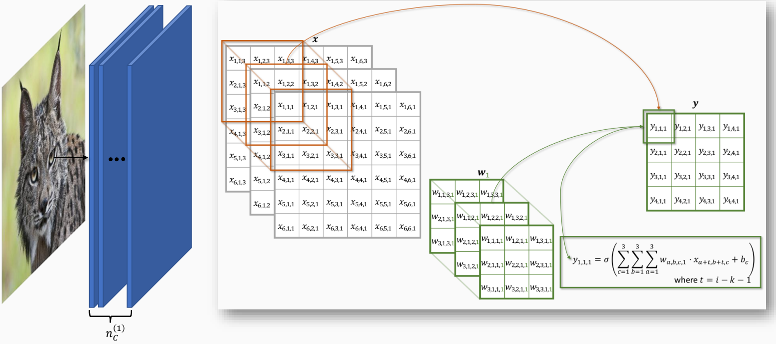
Convolutional Neural Network (CNN)



Convolutional Neural Network (CNN)

Feature extraction

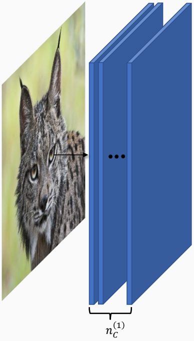
Convolution



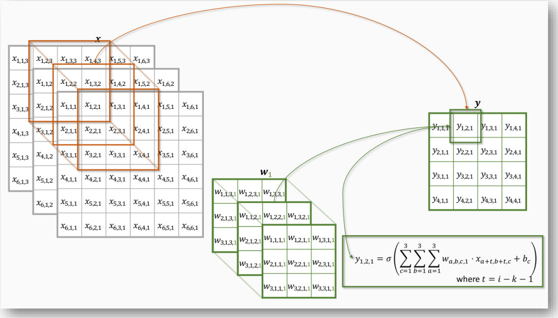
upna

Convolutional Neural Network (CNN)

Feature extraction



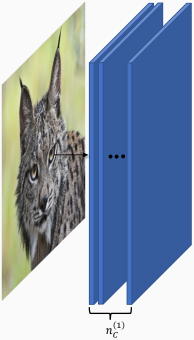
Convolution



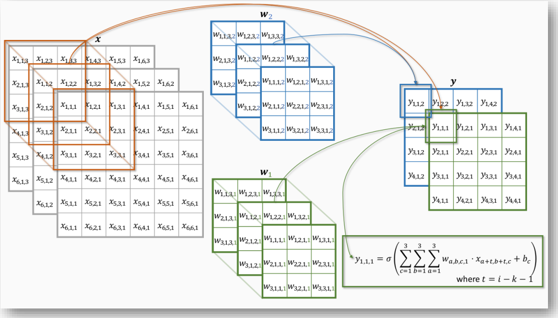
upna

Convolutional Neural Network (CNN)

Feature extraction

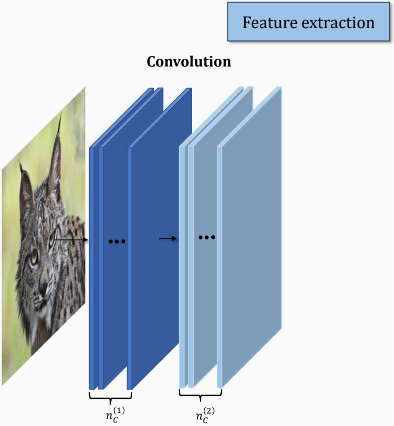


Convolution

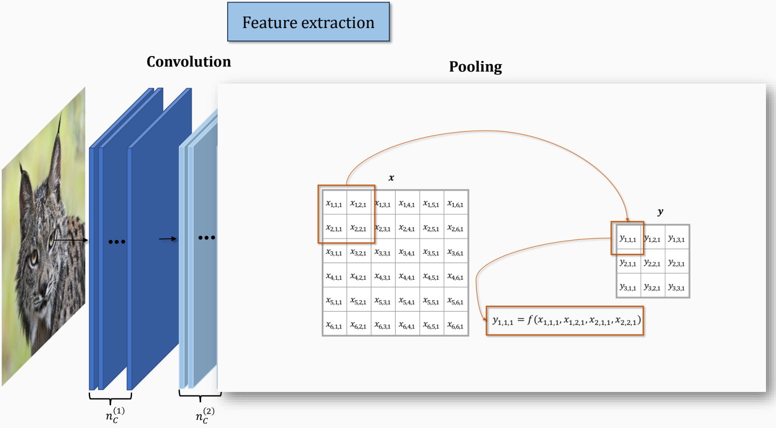


upna

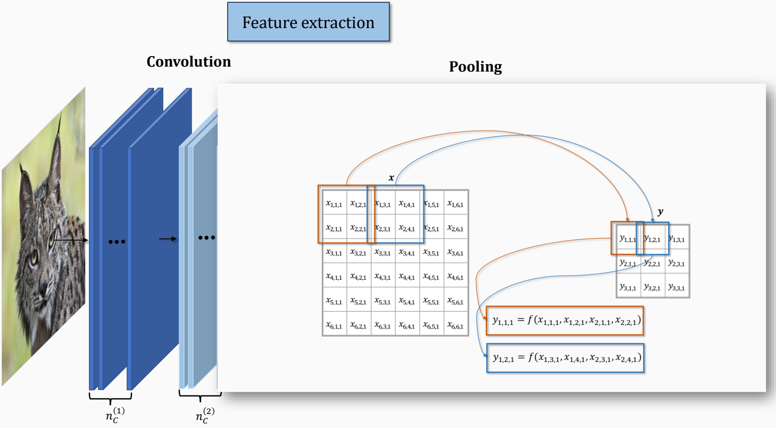
Convolutional Neural Network (CNN)



Convolutional Neural Network (CNN)

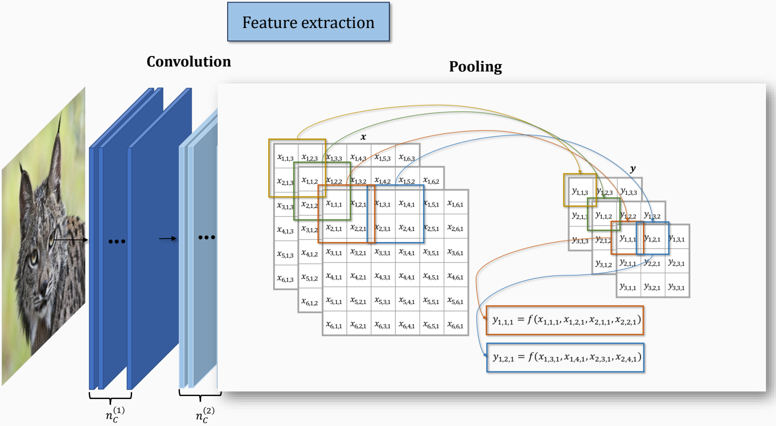


Convolutional Neural Network (CNN)



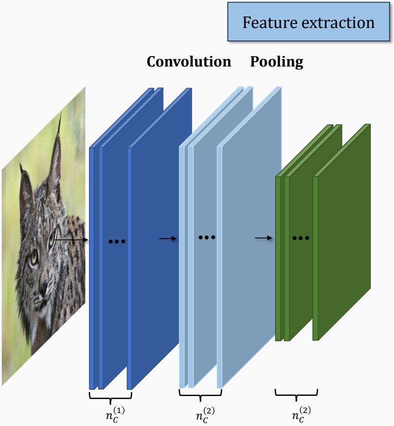
upna

Convolutional Neural Network (CNN)

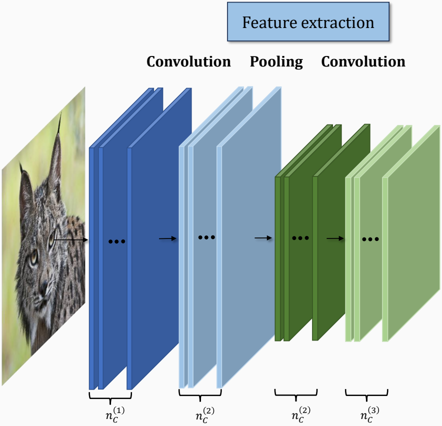


upna

Convolutional Neural Network (CNN)

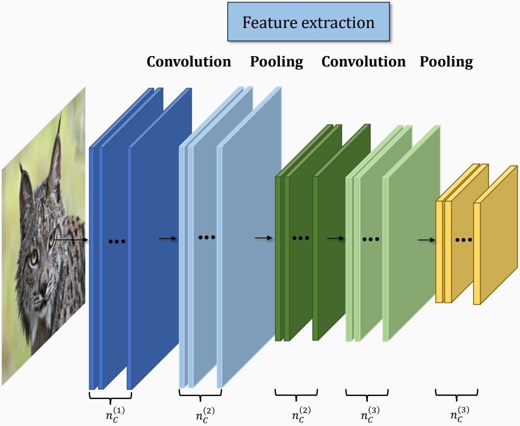


Convolutional Neural Network (CNN)

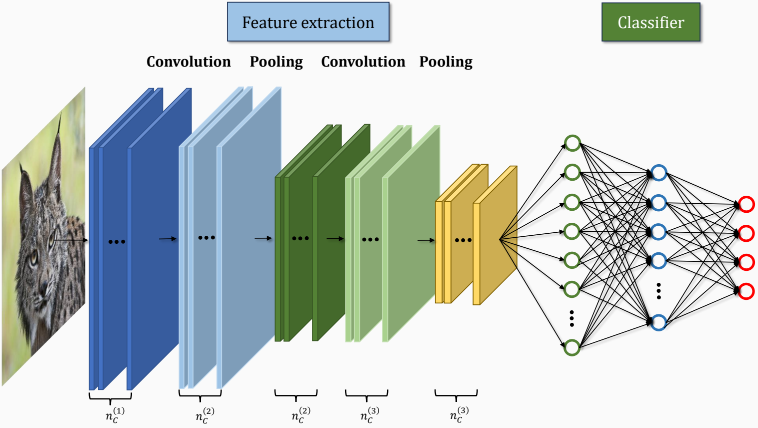


upna

Convolutional Neural Network (CNN)



Convolutional Neural Network (CNN)



upna

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

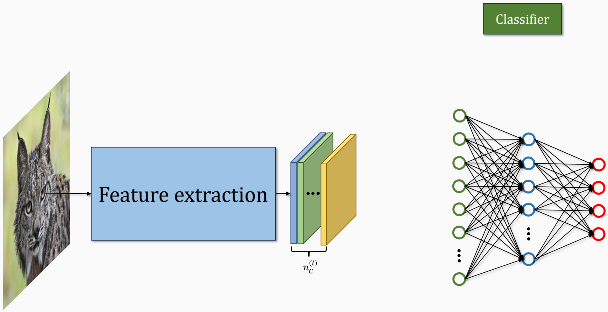
- Summarize features into vectorial form:

upna

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Summarize features into vectorial form:

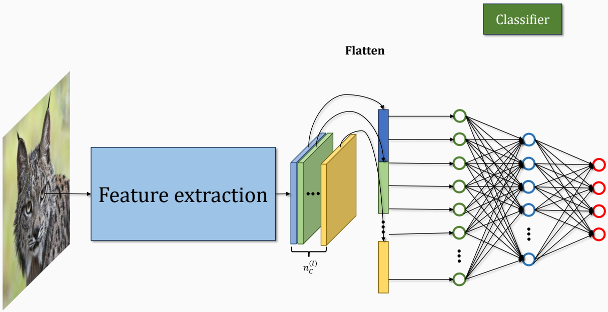


upna

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Summarize features into vectorial form:

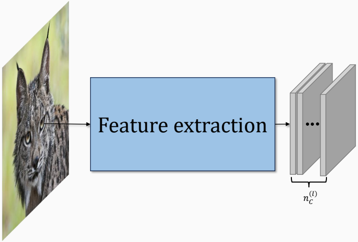


upna

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Summarize features into vectorial form:

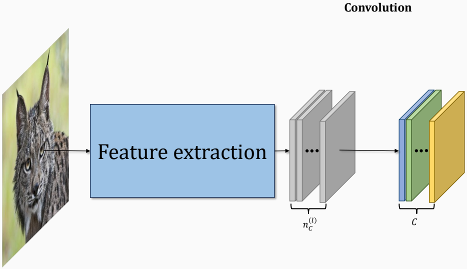


Lin, M. (2013). Network in network. *International Conference on Learning Representations (ICLR), 2013*

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Summarize features into vectorial form:

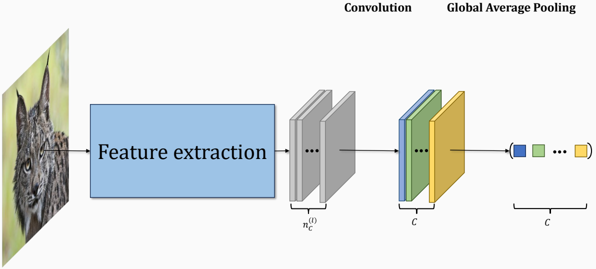


Lin, M. (2013). Network in network. *International Conference on Learning Representations (ICLR), 2013*

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Summarize features into vectorial form:

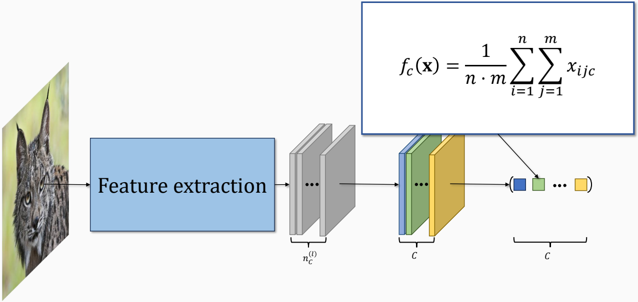


Lin, M. (2013). Network in network. *International Conference on Learning Representations (ICLR), 2013*

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Summarize features into vectorial form:

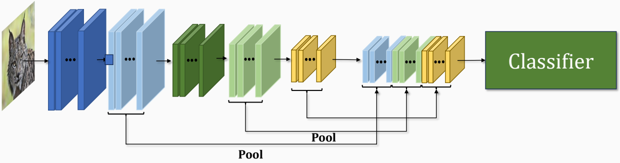


Lin, M. (2013). Network in network. *International Conference on Learning Representations (ICLR), 2013*

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Combining information from different scales:

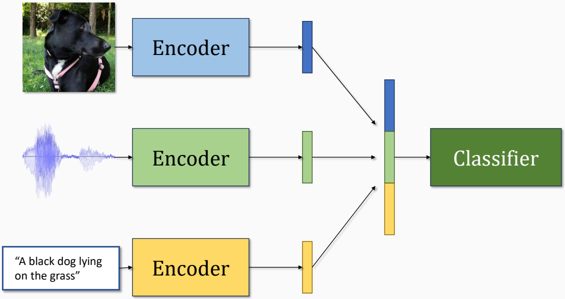


upna

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Combining information from different sources:

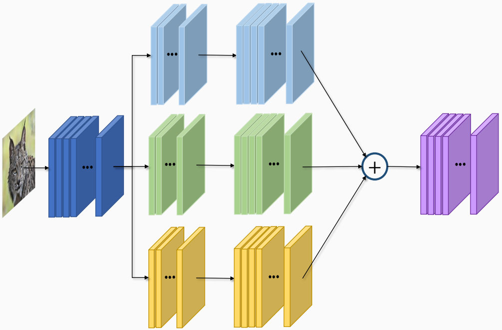


upna

Feature fusion in CNNs

Fusing information is a recurrent problem in NNs:

- Combining information from different “branches”:



Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).

Information fusion

Information Fusion is the process of integrating multiple sources of data to produce a more informed final representation than the one provided by each individual source.

Information fusion

Information Fusion is the process of integrating multiple sources of data to produce a more informed final representation than the one provided by each individual source.

Information fusion covers a wide spectrum of problems:

upna

Information fusion

Information Fusion is the process of integrating multiple sources of data to produce a more informed final representation than the one provided by each individual source.

Information fusion covers a wide spectrum of problems:

- Multi-sensor information fusion

Information fusion

Information Fusion is the process of integrating multiple sources of data to produce a more informed final representation than the one provided by each individual source.

Information fusion covers a wide spectrum of problems:

- Multi-sensor information fusion
- Decision making

Information fusion

Information Fusion is the process of integrating multiple sources of data to produce a more informed final representation than the one provided by each individual source.

Information fusion covers a wide spectrum of problems:

- Multi-sensor information fusion
- Decision making
- Deep Learning (as just illustrated)

Information fusion

Information Fusion is the process of integrating multiple sources of data to produce a more informed final representation than the one provided by each individual source.

Information fusion covers a wide spectrum of problems:

- Multi-sensor information fusion
- Decision making
- Deep Learning (as just illustrated)
- ...

Fusion functions

One of the most recurrent problems is the need to replace a set of values by a single individual representative.

Fusion functions

One of the most recurrent problems is the need to replace a set of values by a single individual representative.

Definition

Any arbitrary function of the type $F : [a, b]^n \to [a, b]$, with $a, b \in \mathbb{R}$ and $a < b$ is called a fusion function.

upna

Fusion functions

One of the most recurrent problems is the need to replace a set of values by a single individual representative.

Definition

Any arbitrary function of the type $F : [a, b]^n \to [a, b]$, with $a, b \in \mathbb{R}$ and $a < b$ is called a fusion function.

Definition

A function $A : [a, b]^n \to [a, b]$ is an aggregation function if:

- A is increasing
- $A(a, \ldots, a) = a$ and $A(b, \ldots, b) = b$

upna

Aggregation functions for feature fusion

Challenges faced in neural networks:

Aggregation functions for feature fusion

Challenges faced in neural networks:

- Range of features is unbounded (real valued data).

upna

Aggregation functions for feature fusion

Challenges faced in neural networks:

- Range of features is unbounded (real valued data).
  - Aggregation functions require boundary conditions!

Aggregation functions for feature fusion

Challenges faced in neural networks:

- Range of features is unbounded (real valued data).
  - Aggregation functions require boundary conditions!
- We have to choose the best aggregation.

Aggregation functions for feature fusion

Challenges faced in neural networks:

- Range of features is unbounded (real valued data).
  - Aggregation functions require boundary conditions!
- We have to choose the best aggregation.
  - How can we consider interaction among data?

Aggregation functions for feature fusion

Challenges faced in neural networks:

- Range of features is unbounded (real valued data).
  - Aggregation functions require boundary conditions!
- We have to choose the best aggregation.
  - How can we consider interaction among data?
  - Which data should we prioritise?

Motivation and objectives

Motivation

- Surprising lack of contributions from the information fusion domain:

Motivation

- Surprising lack of contributions from the information fusion domain:
  - Nowadays, some noteworthy proposals[1,2,3].

---

1 Zeiler, M. D., & Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks, *1st International Conference on Learning Representations, ICLR 2013*, Scottsdale, United States.

2 Bi, Q., Qin, K., Zhang, H., Xie, J., Li, Z., & Xu, K. (2019). APDC-Net: Attention pooling-based convolutional network for aerial scene classification. *IEEE Geoscience and Remote Sensing Letters*, 17(9), 1603-1607.

3 Kortvelesy, R., Morad, S., & Prorok, A. (2023). Generalised f-mean aggregation for graph neural networks. *Advances in Neural Information Processing Systems*, 36, 34439-34450.

upna

Motivation

- Surprising lack of contributions from the information fusion domain:
    - Nowadays, some noteworthy proposals.
    - Also coming from the aggregation theory field[1,2,3].

[1] Forcen, J. I., Pagola, M., Barrenechea, E., & Bustince, H. (2020). Learning ordered pooling weights in image classification. *Neurocomputing*, 411, 45-53.

[2] Dominguez-Catena, I., Paternain, D., & Galar, M. (2021). A study of OWA operators learned in convolutional neural networks. *Applied Sciences*, 11(16), 7195.

[3] Ferrero-Jaurrieta, M., Takáč, Z., Fernández, J., Horanská, Ľ., Dimuro, G. P., Montes, S., ... & Bustince, H. (2022). VCI-LSTM: Vector Choquet integral-based long short-term memory. *IEEE Transactions on Fuzzy Systems*, 31(7), 2238-2250.

upna

Motivation

- Surprising lack of contributions from the information fusion domain:
  - Nowadays, some noteworthy proposals.
  - Also coming from the aggregation theory field.
- Results of the research were applied to COVID-19 prediction:

Motivation

- Surprising lack of contributions from the information fusion domain:
  - Nowadays, some noteworthy proposals.
  - Also coming from the aggregation theory field.
- Results of the research were applied to COVID-19 prediction:
  - Collaboration with Tracasa, Naitec and the University Hospital of Navarra.

upna

Motivation

- Surprising lack of contributions from the information fusion domain:
  - Nowadays, some noteworthy proposals.
  - Also coming from the aggregation theory field.
- Results of the research were applied to COVID-19 prediction:
  - Collaboration with Tracasa, Naitec and the University Hospital of Navarra.
  - Automatic analysis from chest x-ray scans using CNNs.

Main objective

The main objective of this dissertation is to present new methods for fusing the intermediate features of Convolutional Neural Network architectures in the most efficient way possible.

upna

Specific objectives

We will try to do so, by:

- Considering the coalition between neighbouring values through fuzzy integrals.

upna

Specific objectives

We will try to do so, by:

- Considering the coalition between neighbouring values through fuzzy integrals.
- Prioritising high activation values on feature maps through grouping functions.

upna

Specific objectives

We will try to do so, by:

- Considering the coalition between neighbouring values through fuzzy integrals.
- Prioritising high activation values on feature maps through grouping functions.
- Presenting a strategy to construct new pooling operators by combining different functions in a coherent way.

upna

Specific objectives

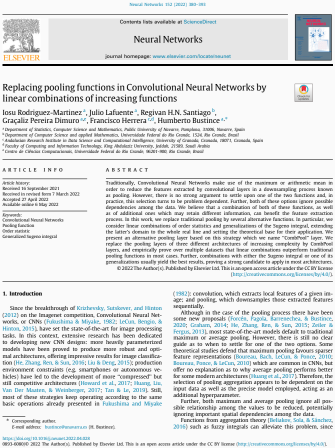
We will try to do so, by:

- Considering the coalition between neighbouring values through fuzzy integrals.
- Prioritising high activation values on feature maps through grouping functions.
- Presenting a strategy to construct new pooling operators by combining different functions in a coherent way.
- Presenting a full CNN pipeline for the detection of COVID-19 positive patients from x-ray scans.

upna

Discussion of research findings

Publication 1



Neural Networks 152 (2022) 380–393

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Replacing pooling functions in Convolutional Neural Networks by linear combinations of increasing functions

Iosu Rodriguez-Martinez [a], Julio Lafuente [a], Regivan H.N. Santiago [b], Graçaliz Pereira Dimuro [a,e], Francisco Herrera [c,d], Humberto Bustince [a,*]

[a] Department of Statistics, Computer Science and Mathematics, Public University of Navarre, Pamplona, 31006, Navarre, Spain
[b] Department of Computer Science and applied Mathematics, Universidade Federal do Rio Grande, 1524, Rio Grande, Brasil
[c] Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, Granada, 18071, Granada, Spain
[d] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[e] Centro de Ciências Computacionais, Universidade Federal do Rio Grande, 96201-900, Rio Grande, Brasil

ARTICLE INFO

Article history:
Received 16 September 2021
Received in revised form 7 March 2022
Accepted 27 April 2022
Available online 6 May 2022

Keywords:
Convolutional Neural Networks
Pooling function
Order statistic
Generalized Sugeno integral

ABSTRACT

Traditionally, Convolutional Neural Networks make use of the maximum or arithmetic mean in order to reduce the features extracted by convolutional layers in a downsampling process known as pooling. However, there is no strong argument to settle upon one of the two functions and, in practice, this selection turns to be problem dependent. Further, both of these options ignore possible dependencies among the data. We believe that a combination of both of these functions, as well as of additional ones which may retain different information, can benefit the feature extraction process. In this work, we replace traditional pooling by several alternative functions. In particular, we consider linear combinations of order statistics and generalizations of the Sugeno integral, extending the latter's domain to the whole real line and setting the theoretical base for their application. We present an alternative pooling layer based on this strategy which we name "CombPool" layer. We replace the pooling layers of three different architectures of increasing complexity by CombPool layers, and empirically prove over multiple datasets that linear combinations outperform traditional pooling functions in most cases. Further, combinations with either the Sugeno integral or one of its generalizations usually yield the best results, proving a strong candidate to apply in most architectures.

1. Introduction

Since the breakthrough of Krizhevsky, Sutskever, and Hinton (2012) on the Imagenet competition, Convolutional Neural Networks, or CNNs (Fukushima & Miyake, 1982; LeCun, Bengio, & Hinton, 2015), have set the state-of-the-art for image processing tasks. In this context, extensive research has been dedicated to developing new CNN designs: more heavily parameterized models have been proved to produce more robust and optimal architectures, offering impressive results for image classification (He, Zhang, Ren, & Sun, 2016; Liu & Deng, 2015); production environment constraints (e.g. smartphones or autonomous vehicles) have led to the development of more "compressed" but still competitive architectures (Howard et al., 2017; Huang, Liu, Van Der Maaten, & Weinberger, 2017; Tan & Le, 2019). Still, most of these strategies keep operating according to the same basic operations already presented in Fukushima and Miyake (1982): convolution, which extracts local features of a given image; and pooling, which downsamples those extracted features sequentially.

Although in the case of the pooling process there have been some new proposals (Forcén, Pagola, Barrenechea, & Bustince, 2020; Graham, 2014; He, Zhang, Ren, & Sun, 2015; Zeiler & Fergus, 2013), most state-of-the-art models default to traditional maximum or average pooling. However, there is still no clear guide as to when to settle for one of the two options. Some theoretical studies defend that maximum pooling favours sparser feature representations (Boureau, Bach, LeCun, & Ponce, 2010; Boureau, Ponce, & LeCun, 2010) which are common in CNNs, but offer no explanation as to why average pooling performs better for some modern architectures (Huang et al., 2017). Therefore, the selection of pooling aggregation appears to be dependent on the input data as well as the precise model employed, acting as an additional hyperparameter.

Further, both maximum and average pooling ignore all possible relationship among the values to be reduced, potentially ignoring important spatial dependencies among the data.

Functions from aggregation theory (Beliakov, Sola, & Sánchez, 2016) such as fuzzy integrals can alleviate this problem, since

* Corresponding author.
E-mail address: bustince@unavarra.es (H. Bustince).

https://doi.org/10.1016/j.neunet.2022.04.028

- **Journal:** Neural Networks
- **Status:** Published.
- **JIF (JCR 2022):** 7.8
- **JCR Ranking Categories:**
  - Computer Science, Artificial Intelligence: 28/145 (Q1)
  - Neurosciences: 26/272 (Q1)

Rodriguez-Martinez, I., Lafuente, J., Santiago, R. H., Dimuro, G. P., Herrera, F., & Bustince, H. (2022). Replacing pooling functions in Convolutional Neural Networks by linear combinations of increasing functions. *Neural Networks*, 152, 380-393.

Combination of pooling operators

- Choosing between max-pooling or avg-pooling is not direct.

[1] C. -Y. Lee, P. Gallagher and Z. Tu (2018), Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 863-875.

Combination of pooling operators

- Choosing between max-pooling or avg-pooling is not direct.
- Combining both options yields better results![1]
    - $f_{mix}(\mathbf{x}) = \alpha \cdot \max_{i=1}^{n} \mathbf{x} + (1-\alpha) \cdot \frac{1}{n} \sum_{i=1}^{n} x_i$, with $\alpha \in [0,1]$

[1] C. -Y. Lee, P. Gallagher and Z. Tu (2018), Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 863-875.

Combination of pooling operators

- Choosing between max-pooling or avg-pooling is not direct.
- Combining both options yields better results![1]
  - $f_{mix}(\mathbf{x}) = \alpha \cdot \max_{i=1}^{n} \mathbf{x} + (1-\alpha) \cdot \frac{1}{n} \sum_{i=1}^{n} x_i$, with $\alpha \in [0,1]$
- The idea can be extended:

[1] C. -Y. Lee, P. Gallagher and Z. Tu (2018), Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 863-875.

Combination of pooling operators

- Choosing between max-pooling or avg-pooling is not direct.
- Combining both options yields better results![1]
  - $f_{mix}(\mathbf{x}) = \alpha \cdot \max_{i=1}^{n} \mathbf{x} + (1-\alpha) \cdot \frac{1}{n} \sum_{i=1}^{n} x_i$, with $\alpha \in [0,1]$
- The idea can be extended: $\rightarrow$ Using other aggregation functions
  - $\{A_i : [a,b]^n \rightarrow [a,b], i \in \{1,...,r\} | A_i$ is increasing and $A_i(\mathbf{a}) = a, A_i(\mathbf{b}) = b\}$
  - $f(\mathbf{x}) = \sum_{i=1}^{r} \alpha_i \cdot \mathbf{A}_i(\mathbf{x})$

[1] C. -Y. Lee, P. Gallagher and Z. Tu (2018), Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 863-875.

Combination of pooling operators

- Choosing between max-pooling or avg-pooling is not direct.
- Combining both options yields better results![1]
    - $f_{mix}(\mathbf{x}) = \alpha \cdot \max_{i=1}^{n} \mathbf{x} + (1-\alpha) \cdot \frac{1}{n} \sum_{i=1}^{n} x_i$, with $\alpha \in [0, 1]$
- The idea can be extended: $\rightarrow$ Using other fusion functions
    - $\{A_i : [a, b]^n \rightarrow [a, b], i \in \{1, ..., r\}\}$
    - $f(\mathbf{x}) = \sum_{i=1}^{r} \alpha_i \cdot \mathbf{A}_i(\mathbf{x})$

[1] C. -Y. Lee, P. Gallagher and Z. Tu (2018), Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 863-875.

upna

Combination of pooling operators

- Choosing between max-pooling or avg-pooling is not direct.
- Combining both options yields better results![1]
  - $f_{mix}(\mathbf{x}) = \alpha \cdot \max_{i=1}^{n} \mathbf{x} + (1-\alpha) \cdot \frac{1}{n} \sum_{i=1}^{n} x_i$, with $\alpha \in [0, 1]$
- The idea can be extended: $\rightarrow$ Using other increasing functions
  - $\{A_i : \mathbb{R}^n \rightarrow \mathbb{R}, i \in \{1, ..., r\} | A_i$ is increasing$\}$
  - $f(\mathbf{x}) = \sum_{i=1}^{r} \alpha_i \cdot \mathbf{A}_i(\mathbf{x})$

[1] C. -Y. Lee, P. Gallagher and Z. Tu (2018), Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 863-875.

Fuzzy integrals

Allow to weigh the coalition among data through a fuzzy measure.

Definition

Let $\mathcal{N} = \{1, \ldots, n\}$. A discrete fuzzy measure on $\mathcal{N}$ is a map $\nu : 2^{\mathcal{N}} \to [0, +\infty)$ such that

- $\nu(\emptyset) = 0$,
- $S \subseteq T \subseteq \mathcal{N}$ implies $\nu(S) \leq \nu(T)$

Fuzzy integrals (II)

Given a fuzzy measure, the Sugeno fuzzy integral is given by:

Definition

The discrete Sugeno integral $S_\nu : \mathbb{R}^n \to \mathbb{R}$ with respect to a fuzzy measure $\nu : 2^{\mathcal{N}} \to [0, +\infty)$ is given by

$$S_\nu(\mathbf{x}) = \max_{i=1,\dots,n} \min\{x_{(i)}, \nu(H_i)\},$$

where $\mathbf{x}_{\nearrow} = (x_{(1)}, x_{(2)}, \dots, x_{(n)})$ is an increasing permutation of $\mathbf{x}$ and $H_i = \{(i), \dots, (n)\}$.

upna

Generalized Sugeno integral

- We also test generalized forms of the Sugeno integral[2].

Definition

Let $\mathcal{N} = \{1, \ldots, n\}$ and let $\mathbb{U}$ be a connected subset of $\mathbb{R}$ such that $0 \in \mathbb{U}$. A $\mathbb{U}$-fuzzy measure on $\mathcal{N}$ is a map $\nu : 2^{\mathcal{N}} \to \mathbb{U}$ such that

- $\nu(\emptyset) = 0$,
- $S \subseteq T \subseteq \mathcal{N}$ implies $\nu(S) \leq \nu(T)$

[2] Bardozzo, F., De La Osa, B., Horanská, Ľ., Fumanal-Idocin, J., delli Priscoli, M., Troiano, L., … & Bustince, H. (2021). Sugeno integral generalization applied to improve adaptive image binarization. Information Fusion, 68, 37-45.

Generalized Sugeno integral

- We also test generalized forms of the Sugeno integral[2].

Definition

Let $\mathbb{U}$ and $\mathbb{I}$ be two connected subsets of $\mathbb{R}$ such that $0 \in \mathbb{U} \subseteq \mathbb{I}$. Let $\nu : 2^{\mathcal{N}} \to \mathbb{U}$ be a $\mathbb{U}$-fuzzy measure. We say that the maps $F : \mathbb{I} \times \mathbb{U} \to \mathbb{I}$ and $G : \mathbb{I}^n \to \mathbb{U}$ are ν-admissible if the map $\mathbf{A} : \mathbb{I}^n \to \mathbb{I}$ given, for $x_1, \dots, x_n \in \mathbb{I}$, by

$$\mathbf{A}(x_1, \dots, x_n) = G(F(x_{\sigma(1)}, \nu(N_1^\sigma)), \dots, F(x_{\sigma(n)}, \nu(N_n^\sigma))),$$

where $\sigma \in \mathbf{x}_{(\nearrow)}$ and $N_i^\sigma = \{\sigma(i), \dots, \sigma(n)\}$, is well defined. Then we set $\mathbf{A} = \mathbf{A}(F, G, \nu)$ and name it the Sugeno-like (F, G, ν)-function.

[2] Bardozzo, F., De La Osa, B., Horanská, Ľ., Fumanal-Idocin, J., delli Priscoli, M., Troiano, L., … & Bustince, H. (2021). Sugeno integral generalization applied to improve adaptive image binarization. Information Fusion, 68, 37-45.

Generalized Sugeno integral (III)

Example

Using $G(\mathbf{x}) = \sum_{i=1}^{n} x_i$, $F(x, y) = x \cdot y$ and a symmetrical fuzzy measure ν we obtain the Sugeno-like (Π, Σ, ν)-function given by

$$\mathbf{D}_\nu(\mathbf{x}) = \sum_{i=1}^{n} x_{\sigma(i)} \cdot \nu(N_i^\sigma)$$

Combining increasing functions

- We test several functions:

Combining increasing functions

- We test several functions:
  - Maximum and arithmetic mean

Combining increasing functions

- We test several functions:
    - Maximum and arithmetic mean
    - Order statistics

Combining increasing functions

- We test several functions:
  - Maximum and arithmetic mean
  - Order statistics
  - Sugeno integral

Combining increasing functions

- We test several functions:
  - Maximum and arithmetic mean
  - Order statistics
  - Sugeno integral
  - Sugeno-like (F, G, ν)-functions

Combining increasing functions

- We test several functions:
  - Maximum and arithmetic mean
  - Order statistics
  - Sugeno integral
  - Sugeno-like (F, G, ν)-functions
- And we want to combine them

upna

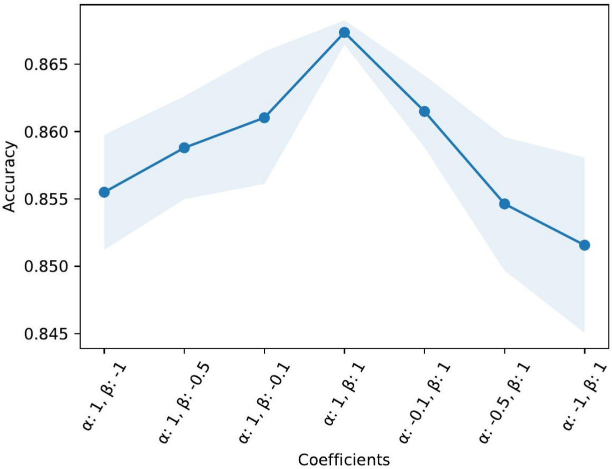
Combining increasing functions

- We test several functions:
    - Maximum and arithmetic mean
    - Order statistics
    - Sugeno integral
    - Sugeno-like (F, G, ν)-functions
- And we want to combine them
    - Preserving monotonicity

Combining increasing functions

- We test several functions:
  - Maximum and arithmetic mean
  - Order statistics
  - Sugeno integral
  - Sugeno-like (F, G, ν)-functions
- And we want to combine them
  - Preserving monotonicity
  - Certain restrictions must be enforced

But is monotonicity (increasingness) important anyway?



Enforcing monotonicity

Let $\mathbf{A}_1, \dots, \mathbf{A}_r : \mathbb{R}^n \to \mathbb{R}$ be increasing functions. We denote $\mathcal{J}(\mathbf{A}_1, \dots, \mathbf{A}_r) = \{(\alpha_1, \dots, \alpha_r) \in \mathbb{R}^n | \sum_{i=1}^{r} \alpha_i \mathbf{A}_i : \mathbb{R}^n \to \mathbb{R}$ is an increasing function$\}$

upna

Enforcing monotonicity

Let $\mathbf{A}_1, \ldots, \mathbf{A}_r : \mathbb{R}^n \rightarrow \mathbb{R}$ be increasing functions. We denote $\mathcal{I}(\mathbf{A}_1, \ldots, \mathbf{A}_r) = \{(\alpha_1, \ldots, \alpha_r) \in \mathbb{R}^n | \sum_{i=1}^{r} \alpha_i \mathbf{A}_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is an increasing function$\}$

Combination of Order Statistics and the Arithmetic Mean

Consider $i_1, \ldots, i_r \in 1, \ldots, n, i_1 < \cdots < i_r, r < n$. Then, for all order statistics $\mathbf{OS}_{i_1}, \ldots, \mathbf{OS}_{i_r}$, it holds that

$$\mathcal{I}(\mathsf{AM}, \mathsf{OS}_{i_1}, \ldots, \mathsf{OS}_{i_r}) = \{(\alpha, \beta_1, \ldots, \beta_r) | \alpha, \alpha + n\beta_1, \ldots, \alpha + n\beta_r \geq 0\}.$$

upna

Enforcing monotonicity (II)

Combination of Order Statistics and the Sugeno integral

Let $\nu : 2^{\mathcal{N}} \to [0, +\infty)$ be a fuzzy measure. If $\alpha_1, \dots, \alpha_n, \alpha_n + \beta \geq 0$, then for all order statistics $\mathsf{OS}_{i_1}, \dots, \mathsf{OS}_{i_r}$ and Sugeno integral S_ν, $\alpha_1 \mathsf{OS}_1 + \dots + \alpha_n \mathsf{OS}_n$ is increasing. If S_ν, $\alpha_1 \mathsf{OS}_1 + \dots + \alpha_n \mathsf{OS}_n$ is increasing and ν is strict in $k \in \mathcal{N}$, then $\alpha_k + \beta \geq 0$; hence if ν is strict, we have that

$$\mathcal{I}(\mathsf{OS}_1, \dots, \mathsf{OS}_n, \mathsf{S}_\nu) = \{(\alpha_1, \dots, \alpha_n, \beta) | \alpha, \alpha + n\beta_1, \dots, \alpha + n\beta_r \geq 0\}.$$

upna

Enforcing monotonicity (II)

Combination of Order Statistics and the Sugeno integral

Let $\nu : 2^{\mathcal{N}} \to [0, +\infty)$ be a fuzzy measure. If $\alpha_1, \dots, \alpha_n, \alpha_n + \beta \geq 0$, then for all order statistics $\mathsf{OS}_{i_1}, \dots, \mathsf{OS}_{i_r}$ and Sugeno integral S_ν, $\alpha_1 \mathsf{OS}_1 + \cdots + \alpha_n \mathsf{OS}_n$ is increasing. If S_ν, $\alpha_1 \mathsf{OS}_1 + \cdots + \alpha_n \mathsf{OS}_n$ is increasing and ν is strict in $k \in \mathcal{N}$, then $\alpha_k + \beta \geq 0$; hence if ν is strict, we have that

$$\mathcal{I}(\mathsf{OS}_1, \dots, \mathsf{OS}_n, \mathsf{S}_\nu) = \{(\alpha_1, \dots, \alpha_n, \beta) | \alpha, \alpha + n\beta_1, \dots, \alpha + n\beta_r \geq 0\}.$$

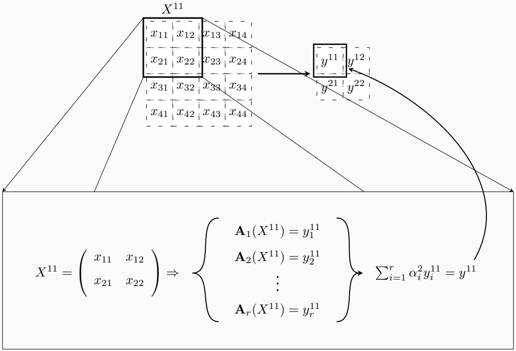
Combination of the Arithmetic Mean and the Sugeno integral

Let $\nu : 2^{\mathcal{N}} \to [0, +\infty)$ be a fuzzy measure. We have

$$\mathcal{I}(\mathsf{AM}, \mathsf{S}_\nu) = \{(\alpha, \beta) | \alpha, \alpha, \alpha + n\beta \geq 0\}.$$

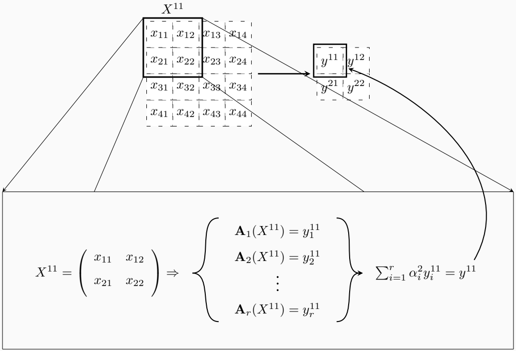
Enforcing monotonicity (III)

We can guarantee monotonicity for all possible combinations by learning positive coefficients:



Enforcing monotonicity (III)

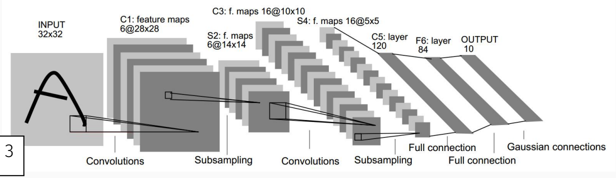
We can guarantee monotonicity for all possible combinations by learning positive coefficients:



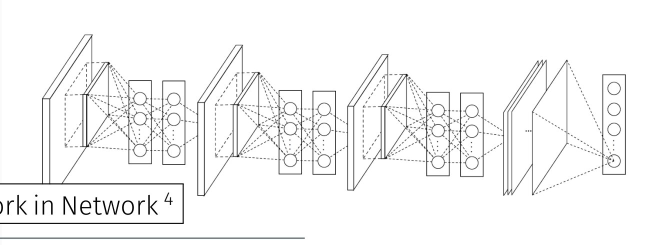
We name the proposal CombPool layers

upna

Experimental framework: Models



LeNet-5 [3]



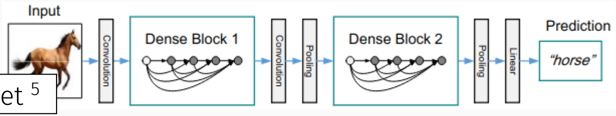
Network in Network [4]

[3] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) (1998), 2278-2324.

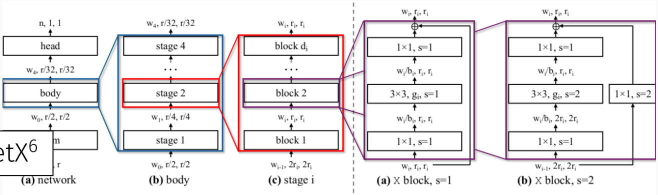
[4] Lin, M., Chen, Q. and Yan, S. (2014) Network in Network. *2nd International Conference on Learning Representations, ICLR 2014*, Banff, AB, 14-16 April 2014.

Experimental framework: Models

DenseNet [5]



RegNetX[6]



---

[5] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[6] Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., & Dollár, P. (2020). Designing network design spaces. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10428-10436).

upna

Experimental framework: Datasets

Dataset	Train	Test	Classes	Colour	Description
MNIST	60000	10000	10	No	Digits from 0 to 9
Fashion MNIST	60000	10000	10	No	Clothing categories
Balanced EMNIST	112800	18800	47	No	Digits and characters
CIFAR10	50000	10000	10	Yes	Real life images
CIFAR100	50000	10000	100	Yes	Real life images

Experimental results

	FASHION dataset		
	LeNet-5	NiN	DenseNet
1st Best	AM	Min	$\mathbf{D}_\nu$ + Min + Max + Median
Accuracy	93.24	93.03	93.79
2nd Best	$\mathbf{S}_\nu$ + AM	Max	AM
Accuracy	93.21	92.99	93.79
3rd Best	$\mathbf{D}_\nu$	$\mathbf{S}_\nu$	AM + Max
Accuracy	93.05	92.67	93.63
	EMNIST dataset		
	LeNet-5	NiN	DenseNet
1st	AM	$\mathbf{D}_\nu$	$\mathbf{S}_\nu$ + AM
Accuracy	87.58	89.27	90.03
2nd	Min + Max + Median	Max	$\mathbf{D}_\nu$ + Max
Accuracy	87.52	89.11	89.97
3rd	$\mathbf{D}_\nu$ + Min + Max + Median	$\mathbf{D}_\nu$ + Max	$\mathbf{D}_\nu$ + Min + Max + Median
Accuracy	87.46	89.09	89.85

Experimental results (II)

	CIFAR10 dataset		
	LeNet-5 1	NiN	DenseNet
1st	$\mathbf{D}_\nu$ + Min + Max	$\mathbf{D}_\nu$	$\mathbf{D}_\nu$ + AM
Accuracy	77.81	88.70	89.87
2nd	Max	$\mathbf{D}_\nu$ + Min + Max	$\mathbf{D}_\nu$ + Min + Max
Accuracy	77.39	88.61	89.83
3rd	$\mathbf{S}_\nu$ + Min + Max	$\mathbf{D}_\nu$ + Min	AM + Min + Max + Median
Accuracy	77.30	88.51	89.83
	CIFAR100 dataset		
	LeNet-5 1	NiN	DenseNet
1st	AM	Max	AM
Accuracy	46.55	57.58	70.78
2nd	$\mathbf{S}_\nu$ + AM	$\mathbf{S}_\nu$ + Min + Max + Median	$\mathbf{D}_\nu$ + Min + Max
Accuracy	46.46	56.08	70.31
3rd	$\mathbf{S}_\nu$ + Max	AM + Min + Max + Median	Min + Max + Median
Accuracy	46.37	55.98	70.21

Other coefficient learning strategies

- Direct optimization of mixing coefficients can be improved:

upna

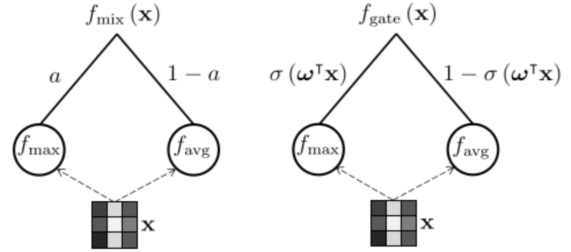
Other coefficient learning strategies

- Direct optimization of mixing coefficients can be improved:
  - CombPool layers are agnostic to coefficient optimization

upna

Other coefficient learning strategies

- Direct optimization of mixing coefficients can be improved:
  - CombPool layers are agnostic to coefficient optimization
  - e. g. using Gated CombPool layers[7]



[7] C. Y. Lee, P. Gallagher and Z. Tu (2018), Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 863-875.

Other coefficient learning strategies

- Direct optimization of mixing coefficients can be improved:
  - CombPool layers are agnostic to coefficient optimization
  - e. g. using Gated CombPool layers

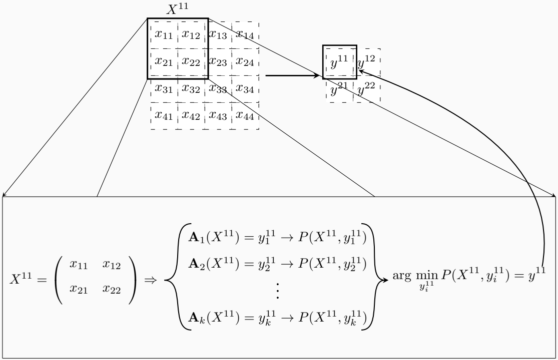
Table 1: Accuracy rate for DenseNet-101 over CIFAR10 dataset

Method	Accuracy
Mixed AM + Max	86.99
Mixed $\mathbf{D}_\nu$ + AM	**89.87**
Gated AM + Max	90.41
Gated $\mathbf{D}_\nu$ + AM	**90.89**

upna

Other coefficient learning strategies

- Direct optimization of mixing coefficients can be improved:
    - CombPool layers are agnostic to coefficient optimization
    - e. g. choosing the best function through penalty-based functions

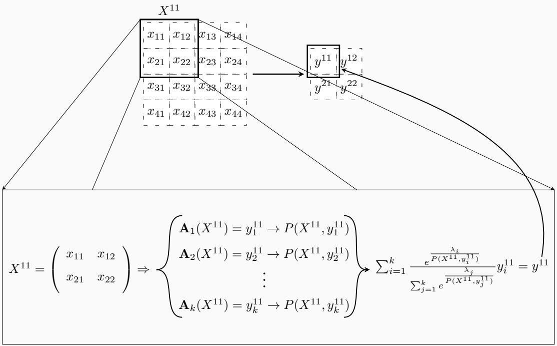


Bustince, H., Beliakov, G., Dimuro, G. P., Bedregal, B., & Mesiar, R. (2017). On the definition of penalty functions in data aggregation. *Fuzzy Sets and Systems*, 323, 1-18.

upna

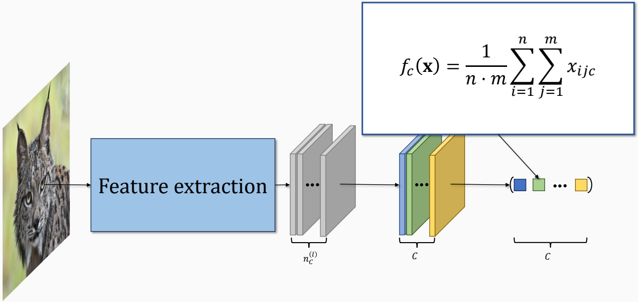
Other coefficient learning strategies

- Direct optimization of mixing coefficients can be improved:
  - CombPool layers are agnostic to coefficient optimization
  - e. g. combining functions according to penalty-based functions



CombPool layers for Global Pooling

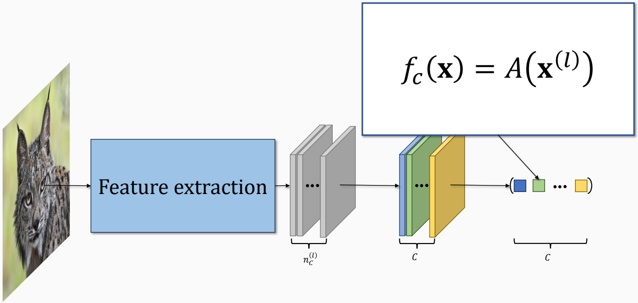
We also replace Global Average Pooling by Global CombPool layers



upna

CombPool layers for Global Pooling

We also replace Global Average Pooling by Global CombPool layers



upna

Experimental results: global CombPool layers

	CIFAR-10			CIFAR-100		
	NiN	DenseNet	RegNetX	NiN	DenseNet	RegNetX
AM	86.11	91.08	94.13	57.16	**70.97**	**74.95**
$Max + AM$	**86.97**	**91.29**	93.77	57.23	68.79	71.43
$Max + \mathbf{S}_\nu$	83.04	91.28	93.40	50.96	65.40	58.96
$Max + \mathbf{D}_\nu$	85.99	90.26	93.27	52.85	66.54	66.66
$AM + \mathbf{S}_\nu$	86.58	91.00	**94.25**	**57.72**	69.68	74.30
$AM + \mathbf{D}_\nu$	86.33	91.08	93.51	52.71	69.10	71.43

Summary of paper 1

- CombPool layers are a solid strategy to combine different reductions
  - Better effect in more complex models
- Global pooling benefits from including the arithmetic mean
  - Test averaging functions (e. g. Moderate Deviation functions).
- $\mathbf{D}_\nu$ offers competitive results
  - Learn aggregation functions from affine transformations[7]

[7] de Hierro, A. F. R. L., Roldán, C., Bustince, H., Fernández, J., Rodriguez, I., Fardoun, H., & Lafuente, J. (2021). Affine construction methodology of aggregation functions. *Fuzzy Sets and Systems*, 414, 146-164.

upna

Publication 2

Information Fusion 99 (2023) 101893

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Generalizing max pooling via (a, b)-grouping functions for Convolutional Neural Networks

Iosu Rodriguez-Martinez [a], Tiago da Cruz Asmus [b], Graçaliz Pereira Dimuro [a,c], Francisco Herrera [d], Zdenko Takáč [e], Humberto Bustince [a,*]

[a] Universidad Pública de Navarra, Departamento de Estadística, Informática y Matemáticas, Campus Arrosadia s/n, Pamplona, 31006, Navarra, Spain
[b] Universidade Federal do Rio Grande, Instituto de Matemática, Estadística e Física, Av. Itália km 08, Campus Carreiros, Rio Grande, 96201-900, Rio Grande, Brazil
[c] Universidade Federal do Rio Grande, Centro de ciencia Computacionais, Av. Itália km 08, Campus Carreiros, Rio Grande, 96201-900, Rio Grande, Brazil
[d] Universidad de Granada, Instituto Andaluz Interuniversitario en Data Science and Computational Intelligence, Av. del Conocimiento 41, Granada, 18071, Granada, Spain
[e] Slovak University of Technology in Bratislava, Faculty of Materials Science and Technology in Trnava, Jana Bottu 2781/25, Trnava, 917 24, Trnava, Slovakia

ARTICLE INFO

Keywords:
Convolutional neural networks
Grouping functions
Pooling functions
Image classification

ABSTRACT

Due to their high adaptability to varied settings and effective optimization algorithm, Convolutional Neural Networks (CNNs) have set the state-of-the-art on image processing jobs for the previous decade. CNNs work in a sequential fashion, alternating between extracting significant features from an input image and aggregating these features locally through "pooling" functions, in order to produce a more compact representation.

Functions like the arithmetic mean or, more typically, the maximum are commonly used to perform this downsampling operation. Despite the fact that many studies have been devoted to the development of alternative pooling algorithms, in practice, "max-pooling" still equals or exceeds most of these possibilities, and has become the standard for CNN construction.

In this paper we focus on the properties that make the maximum such an efficient solution in the context of CNN feature downsampling and propose its replacement by grouping functions, a family of functions that share those desirable properties. In order to adapt these functions to the context of CNNs, we present (a, b)-grouping functions, an extension of grouping functions to work with real valued data. We present different construction methods for (a, b)-grouping functions, and demonstrate their empirical applicability for replacing max-pooling by using them to replace the pooling function of many well-known CNN architectures, finding promising results.

1. Introduction

The irruption of Deep Learning [1] during the last decade has revolutionized the field of machine learning research, with impressive results in fields as diverse as medicine [2,3], natural language processing [4] or synthetic image generation [5]. In the field of computer vision, Convolutional Neural Networks (CNNs) have been established as the state-of-the-art technique for classification [6–8] and segmentation tasks [2,9], among others [10–12].

Unlike traditional Computer Vision methods such as Bag of Features (BoF) [13], the parameters of these models are automatically optimized through gradient descent optimization in a supervised way, easing their application and motivating their wide adoption. CNNs extract complex visual features in a sequential process, generating feature vectors which can be later fed to different algorithms.

If no feature reduction technique were applied, the dimensionality of these feature vectors would be too high. Pooling layers take care of this, performing image downsampling through the fusion of local areas of the feature images the model works with, while trying to preserve the most discriminative values. This data fusion process is usually performed by simple operations such as the arithmetic mean or, more commonly in practice, the maximum. Both theoretical studies [14,15] as well as empirical claims seem to set maximum pooling as the default pooling operator.

Even so, there are some problems with maximum pooling. While providing some amount of shift invariance to the model, maximum

* Corresponding author.
E-mail address: bustince@unavarra.es (H. Bustince).

https://doi.org/10.1016/j.inffus.2023.101893
Received 28 February 2023; Received in revised form 13 June 2023; Accepted 15 June 2023
Available online 21 June 2023

- **Journal:** Information Fusion
- **Status:** Published.
- **JIF (JCR 2023):** 14.7
- **JCR Ranking Categories:**
  - Computer Science, Artificial Intelligence: 4/197 (Q1)
  - Computer Science, Theory & Methods: 2/143 (Q1)

Rodriguez-Martinez, I., da Cruz Asmus, T., Dimuro, G. P., Herrera, F., Takáč, Z., & Bustince, H. (2023). Generalizing max pooling via (a, b)-grouping functions for Convolutional Neural Networks. *Information Fusion*, 99, 101893.

Motivation

- In practice, the maximum is a more common pooling operator than the arithmetic mean

[8] da Cruz Asmus, T., Dimuro, G. P., Bedregal, B., Sanz, J. A., Fernandez, J., Rodriguez-Martinez, I., Mesiar, R., & Bustince, H. (2022). A constructive framework to define fusion functions with floating domains in arbitrary closed real intervals. *Information Sciences*, 610, 800-829.

Motivation

- In practice, the maximum is a more common pooling operator than the arithmetic mean
  - Most of the aggregated information is ignored

[8] da Cruz Asmus, T., Dimuro, G. P., Bedregal, B., Sanz, J. A., Fernandez, J., Rodriguez-Martinez, I., Mesiar, R., & Bustince, H. (2022). A constructive framework to define fusion functions with floating domains in arbitrary closed real intervals. *Information Sciences*, 610, 800-829.

Motivation

- In practice, the maximum is a more common pooling operator than the arithmetic mean
  - Most of the aggregated information is ignored
  - Highest activations are preserved!

[8] da Cruz Asmus, T., Dimuro, G. P., Bedregal, B., Sanz, J. A., Fernandez, J., Rodriguez-Martinez, I., Mesiar, R., & Bustince, H. (2022). A constructive framework to define fusion functions with floating domains in arbitrary closed real intervals. *Information Sciences*, 610, 800-829.

Motivation

- In practice, the maximum is a more common pooling operator than the arithmetic mean
    - Most of the aggregated information is ignored
    - Highest activations are preserved!
- Hypothesis: The "disjunctive" behaviour is the key

[8] da Cruz Asmus, T., Dimuro, G. P., Bedregal, B., Sanz, J. A., Fernandez, J., Rodriguez-Martinez, I., Mesiar, R., & Bustince, H. (2022). A constructive framework to define fusion functions with floating domains in arbitrary closed real intervals. *Information Sciences*, 610, 800-829.

Motivation

- In practice, the maximum is a more common pooling operator than the arithmetic mean
  - Most of the aggregated information is ignored
  - Highest activations are preserved!
- Hypothesis: The “disjunctive” behaviour is the key
  - Multiple families of aggregations with this behaviour: t-conorms, grouping functions...

[8] da Cruz Asmus, T., Dimuro, G. P., Bedregal, B., Sanz, J. A., Fernandez, J., Rodriguez-Martinez, I., Mesiar, R., & Bustince, H. (2022). A constructive framework to define fusion functions with floating domains in arbitrary closed real intervals. *Information Sciences*, 610, 800-829.

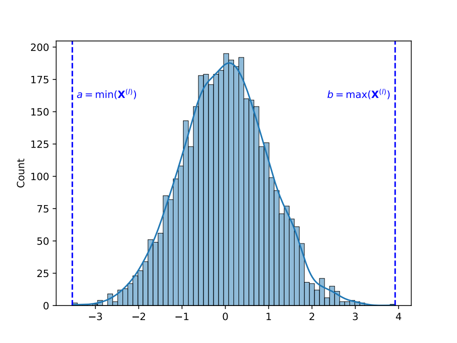
Motivation

- In practice, the maximum is a more common pooling operator than the arithmetic mean
    - Most of the aggregated information is ignored
    - Highest activations are preserved!
- Hypothesis: The "disjunctive" behaviour is the key
    - Multiple families of aggregations with this behaviour: t-conorms, grouping functions...
    - Preliminary promising results using t-conorms[8]

[8] da Cruz Asmus, T., Dimuro, G. P., Bedregal, B., Sanz, J. A., Fernandez, J., Rodriguez-Martinez, I., Mesiar, R., & Bustince, H. (2022). A constructive framework to define fusion functions with floating domains in arbitrary closed real intervals. *Information Sciences*, 610, 800-829.

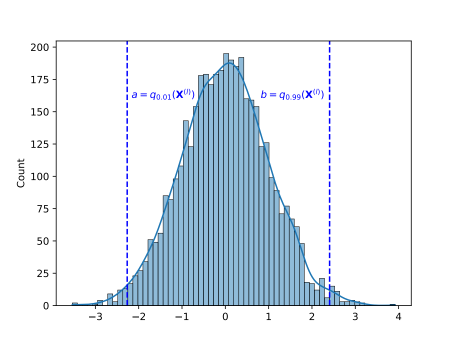
(a, b)-aggregation functions

- In practice, we can usually restrict the range of values to an interval $[a, b]$, with $a < b \in \mathbb{R}$



(a, b)-aggregation functions

- In practice, we can usually restrict the range of values to an interval $[a, b]$, with $a < b \in \mathbb{R}$



- We can apply any fusion/aggregation function!

upna

Restricting the range of values

- Ensuring properties of aggregation functions are preserved in $[a, b]$ is important!

Restricting the range of values

- Ensuring properties of aggregation functions are preserved in $[a, b]$ is important!

Overlap functions in $[0, 1]$[a]

A function $O : [0, 1]^n \to [0, 1]$ is said to be an overlap function if, for all $\mathbf{x} \in [0, 1]^n$, the following conditions hold:

1. O is symmetric;
2. $O(\mathbf{x}) = 0 \iff \Pi_{i=1}^{n} x_i = 0;$
3. $O(\mathbf{x}) = 1 \iff \Pi_{i=1}^{n} x_i = 1;$
4. O is increasing;
5. O is continuous;

[a] Bustince, H., Fernandez, J., Mesiar, R., Montero, J., & Orduna, R. (2010). Overlap functions. *Nonlinear Analysis: Theory, Methods & Applications, 72(3-4)*, 1488-1499.

upna

Restricting the range of values

- Ensuring properties of aggregation functions are preserved in $[a, b]$ is important!

Overlap functions in $[a, b]$[a]

A function $O^{(a,b)} : [a, b]^n \to [a, b]$ is said to be an (a, b)-overlap function if, for all $\mathbf{x} \in [a, b]^n$, the following conditions hold:

1. $O^{(a,b)}$ is symmetric;
2. $O^{(a,b)}(\mathbf{x}) = a \iff \exists x_i \in \mathbf{x}$ such that $x_i = a$;
3. $O^{(a,b)}(\mathbf{x}) = b \iff \forall x_i \in \mathbf{x}, x_i = b$;
4. $O^{(a,b)}$ is increasing;
5. $O^{(a,b)}$ is continuous;

[a] da Cruz Asmus, T., Dimuro, G. P., Bedregal, B., Sanz, J. A., Fernandez, J., Rodriguez-Martinez, I., Mesiar, R., & Bustince, H. (2022). A constructive framework to define fusion functions with floating domains in arbitrary closed real intervals. *Information Sciences*, 610, 800-829.

upna

(a, b)-grouping functions

Definition

A function $G^{(a,b)} : [a, b]^n \to [a, b]$ is said to be an (a, b)-grouping function if, for all $\mathbf{x} \in [a, b]^n$, the following conditions hold:

1. $G^{(a,b)}$ is symmetric;
2. $G^{(a,b)}(\mathbf{x}) = a \iff \forall x_i \in \mathbf{x}, x_i = a$;
3. $G^{(a,b)}(\mathbf{x}) = b \iff \exists x_i \in \mathbf{x}$ such that $x_i = b$;
4. $G^{(a,b)}$ is increasing;
5. $G^{(a,b)}$ is continuous;

upna

Construction methods for (a, b)-grouping functions

Note: Not all grouping functions are (a, b)-grouping functions:

- e. g. $G(\mathbf{x}) = (\max(\mathbf{x}))^p$ is a grouping function but not an (a, b)-grouping function

upna

Construction methods for (a, b)-grouping functions

Note: Not all grouping functions are (a, b)-grouping functions:

- e. g. $G(\mathbf{x}) = (\max(\mathbf{x}))^p$ is a grouping function but not an (a, b)-grouping function

Theorem

Given a function $G : [0, 1]^n \to [0, 1]$, an increasing and bijective function $\phi : [a, b] \to [0, 1]$ and an (a, b)-fusion function $G^{a,b} : [a, b]^n \to [a, b]$ given, for all $x_1, \ldots, x_n \in [a, b]$ by

$$G^{a,b}(\mathbf{x}) = \phi^{-1}(G(\phi(x_1), \ldots, \phi(x_n))),$$

Then, $G^{a,b}$ is an n-dimensional (a, b)-grouping function if and only if G is an n-dimensional grouping function.

upna

Construction methods for (a, b)-grouping functions (II)

Given $\mathbf{G}^{a,b} = \{G_1^{a,b}, \dots, G_m^{a,b}\}$ and $GC^{a,b}$, (a, b)-grouping functions, we also have the following constructions:

- Convex combination of (a, b)-grouping functions:

$$AW^{a,b}_{\mathbf{G}^{a,b}}(\mathbf{x}) = w_1 G_1(\mathbf{x}) + \cdots + w_m G_m(\mathbf{x})$$

upna

Construction methods for (a, b)-grouping functions (II)

Given $\mathbf{G}^{a,b} = \{G_1^{a,b}, \dots, G_m^{a,b}\}$ and $GC^{a,b}$, (a, b)-grouping functions, we also have the following constructions:

- Convex combination of (a, b)-grouping functions:

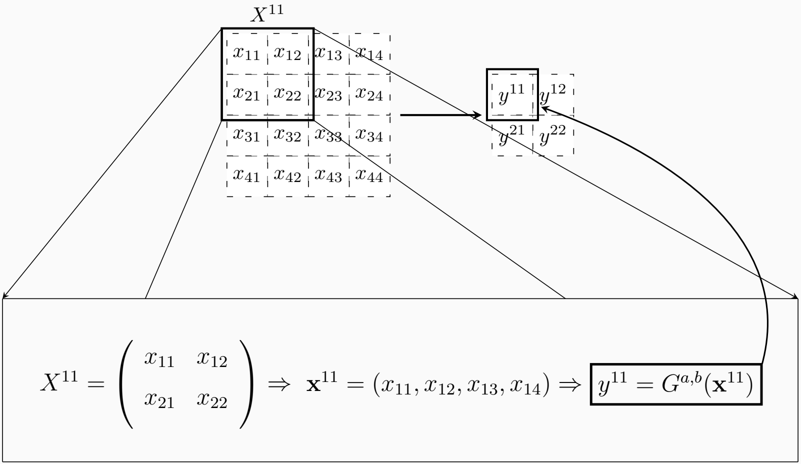
$$AW_{\mathbf{G}^{a,b}}^{a,b}(\mathbf{x}) = w_1 G_1(\mathbf{x}) + \cdots + w_m G_m(\mathbf{x})$$

- Composition of (a, b)-grouping functions:

$$GC_{\mathbf{G}^{a,b}}^{a,b}(\mathbf{x}) = GC^{a,b}\left(G_1(\mathbf{x}), \dots, G_m(\mathbf{x})\right)$$

upna

(a,b)-grouping pooling



upna

Experimental framework: Tested expressions

Name	Core function
$G^{a,b}_{max}$	$G_{max}(\mathbf{x}) = \max_{i=1}^{n} x_i$
$G^{a,b}_{prod}$	$G_{prod}(\mathbf{x}) = 1 - \prod_{i=1}^{n}(1 - x_i)$
$G^{a,b}_{geom}$	$G_{geom}(\mathbf{x}) = 1 - \sqrt[n]{\prod_{i=1}^{n}(1 - x_i)}$
$G^{a,b}_{ob}$	$G_{ob}(\mathbf{x}) = 1 - \sqrt{\min_{i=1}^{n}(1 - x_i) \cdot \prod_{i=1}^{n}(1 - x_i)}$
$G^{a,b}_{u}$	$G_u(\mathbf{x}) = \frac{\max_{i=1}^{n}}{\max_{i=1}^{n} + \sqrt[n]{\prod_{i=1}^{n}(1-x_i)}}$

upna

Experimental framework: results

	CIFAR-10		CIFAR-100	
	VGG16	ResNet	VGG16	ResNet
Avg	**0.915 ± 0.001**	**0.919 ± 0.004**	**$0.682 \pm 0.002/0.891 \pm 0.004$**	$0.681 \pm 0.007/0.902 \pm 0.005$
Max	0.911 ± 0.003	**0.919 ± 0.003**	$0.676 \pm 0.003/0.888 \pm 0.004$	$0.681 \pm 0.005/0.898 \pm 0.004$
$G_{prod}^{a,b}$	0.912 ± 0.003	0.918 ± 0.004	$0.678 \pm 0.004/0.889 \pm 0.004$	$0.664 \pm 0.014/0.891 \pm 0.010$
$G_{ob}^{a,b}$	**0.915 ± 0.002**	0.918 ± 0.002	$0.680 \pm 0.001/0.891 \pm 0.003$	**$0.684 \pm 0.018/0.902 \pm 0.004$**
$AW^{a.b}_{(G_{prod}^{a,b}, G_{ob}^{a,b})}$	0.914 ± 0.002	0.914 ± 0.008	$0.679 \pm 0.002/0.890 \pm 0.001$	**$0.674 \pm 0.016/0.898 \pm 0.009$**
$AW^{a,b}_{(G_{max}^{a,b}, G_{ob}^{a,b})}$	0.914 ± 0.001	**0.923 ± 0.001**	$0.679 \pm 0.004/0.891 \pm 0.002$	$0.671 \pm 0.007/0.898 \pm 0.005$
$G^{a,b}_{max_{(G_{prod}^{a,b}, G_{ob}^{a,b})}}$	0.913 ± 0.001	**0.919 ± 0.004**	$0.678 \pm 0.003/0.888 \pm 0.002$	$0.665 \pm 0.019/0.890 \pm 0.020$
$G^{a,b}_{prod_{(G_{max}^{a,b}, G_{ob}^{a,b})}}$	**0.914 ± 0.001**	0.900 ± 0.016	**$0.681 \pm 0.002/0.889 \pm 0.001$**	**$0.669 \pm 0.027/0.894 \pm 0.006$**

9

upna

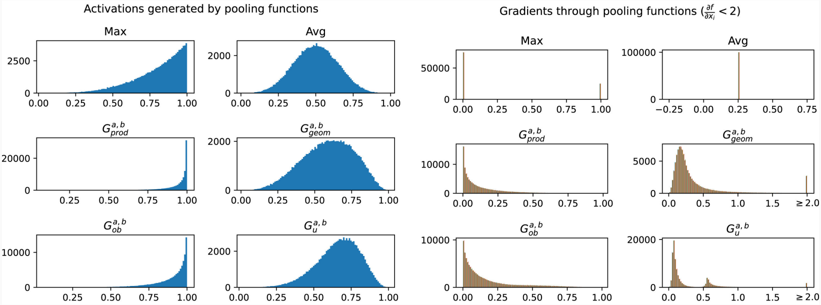
Experimental framework: results

	CIFAR-10		CIFAR-100	
	VGG16	ResNet	VGG16	ResNet
Avg	**0.915 ± 0.001**	**0.919 ± 0.004**	**$0.682 \pm 0.002/0.891 \pm 0.004$**	$0.681 \pm 0.007/0.902 \pm 0.005$
Max	0.911 ± 0.003	**0.919 ± 0.003**	$0.676 \pm 0.003/0.888 \pm 0.004$	$0.681 \pm 0.005/0.898 \pm 0.004$
$G^{a,b}_{prod}$	0.912 ± 0.003	0.918 ± 0.004	$0.678 \pm 0.004/0.889 \pm 0.004$	$0.664 \pm 0.014/0.891 \pm 0.010$
$G^{a,b}_{ob}$	**0.915 ± 0.002**	0.918 ± 0.002	$0.680 \pm 0.001/0.891 \pm 0.003$	**$0.684 \pm 0.018/0.902 \pm 0.004$**
$AW^{a.b}_{(G^{a,b}_{prod},G^{a,b}_{ob})}$	0.914 ± 0.002	0.914 ± 0.008	$0.679 \pm 0.002/0.890 \pm 0.001$	**$0.674 \pm 0.016/0.898 \pm 0.009$**
$AW^{a,b}_{(G^{a,b}_{max},G^{a,b}_{ob})}$	0.914 ± 0.001	**0.923 ± 0.001**	$0.679 \pm 0.004/0.891 \pm 0.002$	$0.671 \pm 0.007/0.898 \pm 0.005$
$G^{a,b}_{max_{(G^{a,b}_{prod},G^{a,b}_{ob})}}$	0.913 ± 0.001	**0.919 ± 0.004**	$0.678 \pm 0.003/0.888 \pm 0.002$	$0.665 \pm 0.019/0.890 \pm 0.020$
$G^{a,b}_{prod_{(G^{a,b}_{max},G^{a,b}_{ob})}}$	**0.914 ± 0.001**	0.900 ± 0.016	**$0.681 \pm 0.002/0.889 \pm 0.001$**	**$0.669 \pm 0.027/0.894 \pm 0.006$**
Mixed pooling	**0.916 ± 0.002**	0.922 ± 0.002	**$0.683 \pm 0.002/0.892 \pm 0.002$**	$0.680 \pm 0.002/0.901 \pm 0.001$
Gated pooling	0.913 ± 0.003	0.922 ± 0.002	$0.682 \pm 0.003/0.892 \pm 0.001$	$0.686 \pm 0.003/0.901 \pm 0.003$
Attention pooling[9]	0.884 ± 0.008	**0.923 ± 0.003**	$0.614 \pm 0.006/0.850 \pm 0.008$	**$0.681 \pm 0.005/0.903 \pm 0.004$**

[9] Bi, Q., Qin, K., Zhang, H., Xie, J., Li, Z., & Xu, K. (2019). APDC-Net: Attention pooling-based convolutional network for aerial scene classification. *IEEE Geoscience and Remote Sensing Letters*, 17(9), 1603-1607.

upna

Effect of (a, b)-grouping functions

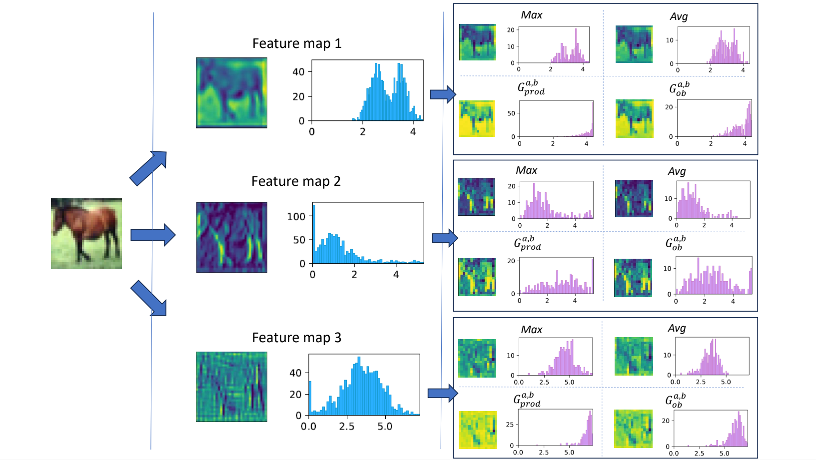


upna

Mitigating exploding gradient with the arithmetic mean

	CIFAR-10		CIFAR-100	
	VGG16	ResNet	VGG16	ResNet
Avg	0.915 ± 0.001	0.919 ± 0.004	$0.682 \pm 0.002/0.891 \pm 0.004$	$0.681 \pm 0.007/0.902 \pm 0.005$
Max	0.911 ± 0.003	0.919 ± 0.003	$0.676 \pm 0.003/0.888 \pm 0.004$	$0.681 \pm 0.005/0.898 \pm 0.004$
Best grouping	$\mathbf{0.916 \pm 0.002}$	$\mathbf{0.923 \pm 0.001}$	$0.681 \pm 0.002/0.889 \pm 0.001$	$0.684 \pm 0.018/0.902 \pm 0.004$
Mixed pooling	$\mathbf{0.916 \pm 0.002}$	0.922 ± 0.002	$0.683 \pm 0.002/0.892 \pm 0.002$	$0.680 \pm 0.002/0.901 \pm 0.001$
$AW^{a,b}_{(Avg,G^{a,b}_{ob})}$	0.914 ± 0.001	0.921 ± 0.002	$\mathbf{0.681 \pm 0.001/0.893 \pm 0.001}$	$\mathbf{0.684 \pm 0.002/0.904 \pm 0.005}$
$AW^{a,b}_{(Avg,G^{a,b}_{prod})}$	0.915 ± 0.001	$\mathbf{0.923 \pm 0.002}$	$0.681 \pm 0.003/0.892 \pm 0.001$	$0.677 \pm 0.012/0.900 \pm 0.006$

Effect of (a, b)-grouping functions (II)



upna

Summary of paper 2

- (a, b)-grouping functions generalize max-pooling
  - While improving gradient flow
- Some expressions can incur in exploding gradient problems
  - Solvable with gradient clipping/mixed pooling
- Competitive with more complex alternatives
  - Requires no additional parameters

Publication 3

Expert Systems With Applications 235 (2024) 121162

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A study on the suitability of different pooling operators for Convolutional Neural Networks in the prediction of COVID-19 through chest x-ray image analysis

Iosu Rodriguez-Martinez [a], Pablo Ursua-Medrano [a], Javier Fernandez [a], Zdenko Takáč [b], Humberto Bustince [a,*]

[a] Universidad Pública de Navarra, Departamento de Estadística, Informática y Matemáticas, Campus Arrosadía s/n, Pamplona, 31006, Navarra, Spain
[b] Slovak University of Technology in Bratislava, Faculty of Materials Science and Technology in Trnava, Jana Bottu 2781/25, Trnava, 917 24, Trnava, Slovakia

ARTICLE INFO

Keywords:
Convolutional Neural Networks
Pooling functions
Aggregation functions
COVID-19
SARS-CoV-2

ABSTRACT

The 2019 coronavirus disease outbreak, caused by the severe acute respiratory syndrome type-2 virus (SARS-CoV-2), was declared a pandemic in March 2020. Since its emergence to the present day, this disease has brought multiple countries to the brink of health care collapse during several waves of the disease. One of the most common tests performed on patients is chest x-ray imaging. These images show the severity of the patient's illness and whether it is indeed covid or another type of pneumonia. Automated assessment of this type of imaging could alleviate the time required for physicians to treat and diagnose each patient. To this end, in this paper we propose the use of Convolutional Neural Networks (CNNs) to carry out this process. The aim of this paper is twofold. Firstly, we present a *pipeline* adapted to this problem, covering all steps from the preprocessing of the datasets to the generation of classification models based on CNNs. Secondly, we have focused our study on the modification of the information fusion processes of this type of architectures, in the *pooling* layers. We propose a number of aggregation theory functions that are suitable to replace classical processes and have shown their benefits in past applications, and study their performance in the context of the x-ray classification problem. We find that replacing the feature reduction processes of CNNs leads to drastically different behaviours of the final model, which can be beneficial when prioritizing certain metrics such as precision or recall.

1. Introduction

The emergence of COVID-19 has posed a tremendous challenge at a health (Bader, Manla, Atallah, & Starling, 2021; Boden et al., 2021; Romero et al., 2022), societal (Singh & Singh, 2020) and economical level (Meyer, Prescott, & Sheng, 2022; Ritter & Pedersen, 2020). Impacts of the COVID-19 pandemic are still being felt nowadays, for example in education (Onyema et al., 2020; Ordorika, 2020), with schools having closed their doors and switched to online learning, which has been a challenge for many students and teachers. Futhermore, many countries have closed their borders and instituted travel restrictions in an effort to control the spread of the virus. This has led to a decrease in international travel and a corresponding decrease in revenue for the tourism industry (Gössling, Scott, & Hall, 2020).

In the initial stages of the pandemic, the identification and triage of the sickness was complex. Despite Computed Tomography (CT) and Reverse Transcription Polymerase Chain Reaction (RT-PCR) being reliable testing mechanisms, economical, temporal and logistical issues made the obtainment of such data complex (Tabik et al., 2020). Chest X-ray (CXR) images, on the other hand, were readily accessible, which propitiated the proliferation of several public datasets to this end (Cohen, Morrison, & Dao, 2020), even if several of them showed class imbalances and biases (Tabik et al., 2020).

Convolutional Neural Networks (CNNs) are a powerful tool for dealing with image classification tasks (He, Zhang, Ren, & Sun, 2016; Huang, Liu, Van Der Maaten, & Weinberger, 2017; Simonyan & Zisserman, 2015). A number of works have seen success to this regard in the context of chest X-ray classification, both for the diagnosis of COVID-19 (Garg, Salehi, La Rocca, Garner, & Duncan, 2022; Li, Zeng, Wu, & Clawson, 2022; Narin, Kaya, & Pamuk, 2021) as well as

* Corresponding author.
E-mail addresses: iosu.rodriguez@unavarra.es (I. Rodriguez-Martinez), pablo.ursua@unavarra.es (P. Ursua-Medrano), fcojavier.fernandez@unavarra.es (J. Fernandez), zdenko.takac@stuba.sk (Z. Takáč), bustince@unavarra.es (H. Bustince).

https://doi.org/10.1016/j.eswa.2023.121162
Received 13 January 2023; Received in revised form 4 May 2023; Accepted 7 August 2023
Available online 16 August 2023

- **Journal:** Expert Systems with Applications
- **Status:** Published.
- **JIF (JCR 2023):** 7.5
- **JCR Ranking Categories:**
  - Computer Science, Artificial Intelligence: 24/197 (Q1)
  - Engineering, Electrical & Electronic: 25/352 (Q1)

Rodriguez-Martinez, I., Ursua-Medrano, P., Fernandez, J., Takáč, Z., & Bustince, H. (2024). A study on the suitability of different pooling operators for Convolutional Neural Networks in the prediction of COVID-19 through chest x-ray image analysis. *Expert Systems with Applications*, 235, 121162.

upna

Motivation

Joint effort between Tracasa Instrumental, Naitec, the University Hospital of Navarre and the Public University of Navarre. Compute power was provided by Nasertic

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

upna

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- Several problems:

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

upna

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- Several problems:
    - Samples taken with different equipment/machinery.

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- Several problems:
  - Samples taken with different equipment/machinery.
  - Mix-up of PosteroAnterior (PA) and AnteroPosterior (AP) scans.

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- Several problems:
  - Samples taken with different equipment/machinery.
  - Mix-up of PosteroAnterior (PA) and AnteroPosterior (AP) scans.
  - Non-reliable annotations.

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

upna

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- Several problems:
  - Samples taken with different equipment/machinery.
  - Mix-up of PosteroAnterior (PA) and AnteroPosterior (AP) scans.
  - Non-reliable annotations.
  - Extreme unbalance with respect to COVID-19 cases.

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- We use the COVIDGR dataset[11].

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

[11] Tabik, S., Gómez-Ríos, A., Martín-Rodríguez, J. L., Sevillano-García, I., Rey-Area, M., Charte, D., ... & Herrera, F. (2020). COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE journal of biomedical and health informatics*, 24(12), 3595-3605.

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- We use the COVIDGR dataset[11].
  - Data collection in collaboration with Hospital Universitario Clínico San Cecilio from Granada.

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

[11] Tabik, S., Gómez-Ríos, A., Martín-Rodríguez, J. L., Sevillano-García, I., Rey-Area, M., Charte, D., ... & Herrera, F. (2020). COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE journal of biomedical and health informatics*, 24(12), 3595-3605.

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- We use the COVIDGR dataset[11].
    - Data collection in collaboration with Hospital Universitario Clínico San Cecilio from Granada.
    - Annotations based on RT-PCR test and expert radiologist evaluation.

---

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

[11] Tabik, S., Gómez-Ríos, A., Martín-Rodríguez, J. L., Sevillano-García, I., Rey-Area, M., Charte, D., ... & Herrera, F. (2020). COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE journal of biomedical and health informatics*, 24(12), 3595-3605.

Dataset

- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- We use the COVIDGR dataset[11].
    - Data collection in collaboration with Hospital Universitario Clínico San Cecilio from Granada.
    - Annotations based on RT-PCR test and expert radiologist evaluation.
    - Homogeneous procedure for scan generation.

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

[11] Tabik, S., Gómez-Ríos, A., Martín-Rodríguez, J. L., Sevillano-García, I., Rey-Area, M., Charte, D., ... & Herrera, F. (2020). COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE journal of biomedical and health informatics*, 24(12), 3595-3605.

Dataset

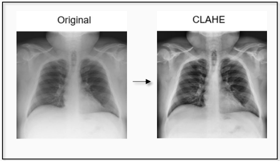
- High proliferation of CXR-datasets during the early steps of the outbreak[10]
- We use the COVIDGR dataset[11].
  - Data collection in collaboration with Hospital Universitario Clínico San Cecilio from Granada.
  - Annotations based on RT-PCR test and expert radiologist evaluation.
  - Homogeneous procedure for scan generation.
  - 852 images: 426 positive / 426 negative.

[10] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint* arXiv:2003.11597.

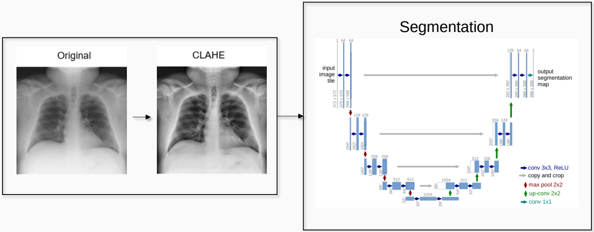
[11] Tabik, S., Gómez-Ríos, A., Martín-Rodríguez, J. L., Sevillano-García, I., Rey-Area, M., Charte, D., … & Herrera, F. (2020). COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE journal of biomedical and health informatics*, 24(12), 3595-3605.

upna

Preprocessing pipeline

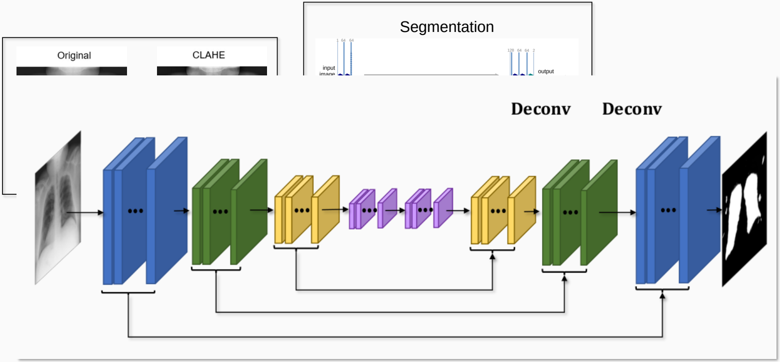


Preprocessing pipeline



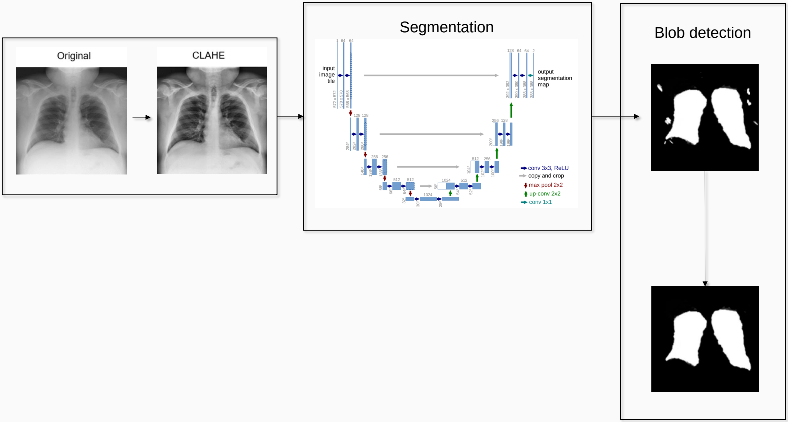
upna

Preprocessing pipeline



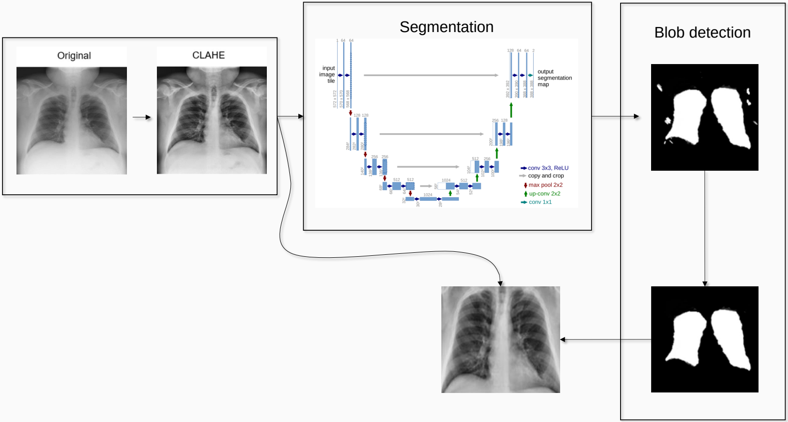
Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015*, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

Preprocessing pipeline

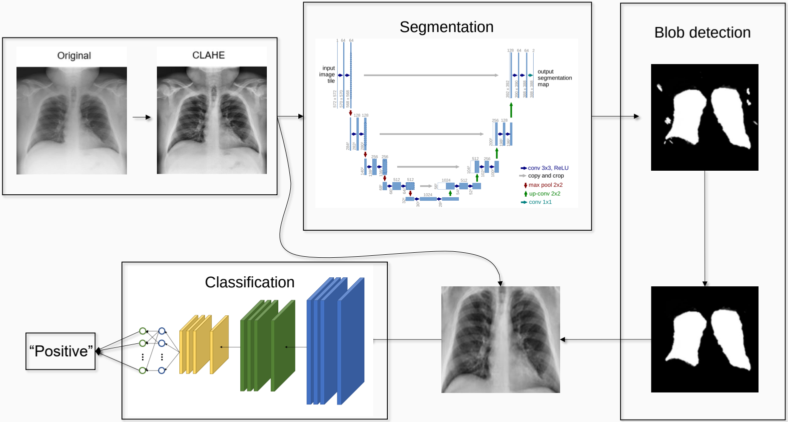


upna

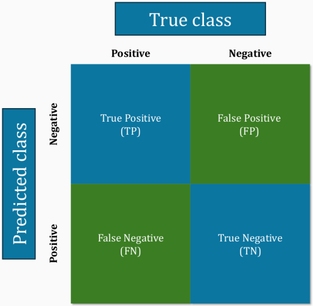
Preprocessing pipeline



Preprocessing pipeline



Evaluation metrics



Evaluation metric	Expression
Accuracy rate	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	$2\,\frac{precision \cdot recall}{precision+recall}$

upna

Experimental framework

- Model: DenseNet-121

Experimental framework

- Model: DenseNet-121
- Pooling layers tested:

upna

Experimental framework

- Model: DenseNet-121
- Pooling layers tested:
  - Max-pool and avg-pool

Experimental framework

- Model: DenseNet-121
- Pooling layers tested:
  - Max-pool and avg-pool
  - Order statistics and sum

upna

Experimental framework

- Model: DenseNet-121
- Pooling layers tested:
  - Max-pool and avg-pool
  - Order statistics and sum
  - Choquet and Sugeno integral

upna

Experimental framework

- Model: DenseNet-121
- Pooling layers tested:
    - Max-pool and avg-pool
    - Order statistics and sum
    - Choquet and Sugeno integral
    - $\mathbf{D}_\nu$ Sugeno-like (Π, Σ, ν)-function

upna

Experimental framework

- Model: DenseNet-121
- Pooling layers tested:
  - Max-pool and avg-pool
  - Order statistics and sum
  - Choquet and Sugeno integral
  - $\mathbf{D}_\nu$ Sugeno-like (Π, Σ, ν)-function
  - (a, b)-grouping functions

Experimental framework

- Model: DenseNet-121
- Pooling layers tested:
  - Max-pool and avg-pool
  - Order statistics and sum
  - Choquet and Sugeno integral
  - $\mathbf{D}_\nu$ Sugeno-like (Π, Σ, ν)-function
  - (a, b)-grouping functions
  - CombPool layers

upna

Experimental framework

- Model: DenseNet-121
- Pooling layers tested:
  - Max-pool and avg-pool
  - Order statistics and sum
  - Choquet and Sugeno integral
  - $\mathbf{D}_\nu$ Sugeno-like (Π, Σ, ν)-function
  - (a, b)-grouping functions
  - CombPool layers
- Metrics report the mean of 5 independent 5-fold cross validated tests per model.

upna

Individual pooling functions

Table 2: Mean results for models which use individual functions.

Pooling Function	Positive class Precision	Recall	F1	Accuracy
AM	0.743 ± 0.038	0.697 ± 0.057	0.717 ± 0.032	0.726 ± 0.026
Max	0.733 ± 0.035	0.678 ± 0.073	0.701 ± 0.044	0.714 ± 0.031
Min	0.729 ± 0.052	0.688 ± 0.064	0.705 ± 0.030	0.712 ± 0.028
Median	0.741 ± 0.035	0.702 ± 0.060	0.705 ± 0.030	$\mathbf{0.727} \pm 0.029$
Sum	0.739 ± 0.042	0.707 ± 0.053	$\mathbf{0.720} \pm 0.026$	0.726 ± 0.024
S_ν	0.732 ± 0.041	0.681 ± 0.076	0.702 ± 0.039	0.713 ± 0.028
D_ν	0.730 ± 0.042	0.693 ± 0.055	0.709 ± 0.036	0.716 ± 0.032
Ch_ν	0.736 ± 0.042	0.680 ± 0.067	0.704 ± 0.041	0.716 ± 0.031

upna

Individual pooling functions

Table 2: Best results of individual runs for models which use individual functions.

Pooling Function	Positive class Precision	Recall	F1	Accuracy
AM	0.752	0.788	0.770	0.764
Max	0.797	0.788	0.792	0.794
Min	0.783	0.764	0.773	0.776
Median	0.786	0.823	**0.804**	**0.800**
Sum	0.736	0.788	0.761	0.752
S_ν	0.699	0.847	0.765	0.741
D_ν	0.755	0.800	0.777	0.770
Ch_ν	0.789	0.705	0.745	0.758

upna

CombPool layers

Table 3: Mean results for models which make use of CombPool layers.

Pooling Function	Positive class			Accuracy
	Precision	Recall	F1	
AM + Max	0.717 ± 0.040	0.710 ± 0.057	0.712 ± 0.038	0.713 ± 0.035
AM + Sum	0.745 ± 0.047	0.705 ± 0.053	$\mathbf{0.722} \pm 0.029$	$\mathbf{0.729} \pm 0.029$
AM + Median	0.745 ± 0.047	0.705 ± 0.053	0.722 ± 0.029	0.716 ± 0.026
D_ν + AM	0.738 ± 0.046	0.707 ± 0.050	0.720 ± 0.031	0.726 ± 0.030
D_ν + Median	0.725 ± 0.042	0.716 ± 0.052	0.719 ± 0.032	0.720 ± 0.032

CombPool layers

Table 3: Best results of individual runs for models which make use of CombPool layers.

Pooling Function	Positive class			Accuracy
	Precision	Recall	F1	
AM + Max	0.766	0.802	0.784	0.779
AM + Sum	0.787	0.741	0.763	0.770
AM + Median	0.777	0.732	0.754	0.761
D_ν + AM	0.767	0.776	0.771	0.770
D_ν + Median	0.790	0.790	**0.790**	**0.790**

upna

(a, b)-grouping functions

Table 4: Mean results for models which use (a, b)-grouping functions.

Pooling Function	Positive class Precision	Recall	F1	Accuracy
$G_{max}^{(a,b)}$	0.758 ± 0.078	0.563 ± 0.132	$\mathbf{0.633} \pm 0.094$	$\mathbf{0.684} \pm 0.05$
$G_{u}^{(a,b)}$	0.801 ± 0.056	0.472 ± 0.059	0.577 ± 0.112	0.673 ± 0.049
$G_{geom}^{(a,b)}$	0.796 ± 0.187	0.375 ± 0.191	0.479 ± 0.220	0.640 ± 0.076
$AW^{(a,b)}_{(G_{max}^{(a,b)}, G_{geom}^{(a,b)})}$	0.808 ± 0.063	0.480 ± 0.148	0.585 ± 0.112	0.678 ± 0.050
$AW^{(a,b)}_{(G_{max}^{(a,b)}, G_{ob}^{(a,b)})}$	0.487 ± 0.086	0.829 ± 0.206	0.607 ± 0.112	0.479 ± 0.116
$AW^{(a,b)}_{(G_{max}^{(a,b)}, G_{ob}^{(a,b)}, G_{prod}^{(a,b)})}$	0.424 ± 0.208	0.695 ± 0.371	0.515 ± 0.250	0.492 ± 0.114
$G_{max}^{(a,b)*}$	0.766 ± 0.078	0.530 ± 0.154	0.610 ± 0.112	0.679 ± 0.053
$G_{geom}^{(a,b)*}$	0.796 ± 0.187	0.375 ± 0.191	0.479 ± 0.221	0.644 ± 0.077
$G_{prod}^{(a,b)*}$	0.446 ± 0.246	0.687 ± 0.364	0.507 ± 0.246	0.490 ± 0.105

(a,b)-grouping functions

Table 4: Best results of individual runs for models which use (a,b)-grouping functions.

Pooling Function	Positive class			Accuracy
	Precision	Recall	F1	
$G^{(a,b)}_{max}$	0.820	0.717	0.767	0.782
$G^{(a,b)}_{u}$	0.880	0.694	0.776	0.800
$G^{(a,b)}_{geom}$	0.514	1.000	0.679	0.529
$AW^{(a,b)}_{(G^{(a,b)}_{max},G^{(a,b)}_{geom})}$	0.835	0.717	0.772	0.788
$AW^{(a,b)}_{(G^{(a,b)}_{max},G^{(a,b)}_{ob})}$	0.615	1.000	0.762	0.641
$AW^{(a,b)}_{(G^{(a,b)}_{max},G^{(a,b)}_{ob},G^{(a,b)}_{prod})}$	0.732	1.000	**0.845**	**0.817**
$G^{(a,b)*}_{max}$	0.780	0.752	0.766	0.770
$G^{(a,b)*}_{geom}$	0.857	0.705	0.774	0.794
$G^{(a,b)*}_{prod}$	1.000	0.624	0.768	0.812

upna

Global Pool layers

Table 5: Mean results for models which replace Global Average Pooling by other aggregation functions or combinations of aggregation functions.

Pooling Function	Positive class Precision	Recall	F1	Accuracy
AM	0.743 ± 0.038	0.697 ± 0.057	**0.717 ± 0.032**	**0.726 ± 0.026**
D_ν	0.892 ± 0.055	0.418 ± 0.127	0.555 ± 0.113	0.680 ± 0.049
Ch_ν	0.891 ± 0.061	0.453 ± 0.123	0.587 ± 0.101	0.694 ± 0.045
$Median$	0.849 ± 0.118	0.393 ± 0.228	0.488 ± 0.213	0.648 ± 0.0785
$AM + Max$	0.830 ± 0.068	0.585 ± 0.117	0.674 ± 0.076	**0.726 ± 0.037**
$AM + Sum$	0.918 ± 0.058	0.383 ± 0.115	0.527 ± 0.101	0.669 ± 0.040
$AM + Median$	0.90 ± 0.059	0.384 ± 0.128	0.524 ± 0.120	0.668 ± 0.048
$D_\nu + AM$	0.895 ± 0.053	0.415 ± 0.154	0.546 ± 0.148	0.679 ± 0.058

Summary of paper 3

- Combining the arithmetic mean with other pooling operators improves performance

upna

Summary of paper 3

- Combining the arithmetic mean with other pooling operators improves performance
  - Effect similar to residual connections!

upna

Summary of paper 3

- Combining the arithmetic mean with other pooling operators improves performance
    - Effect similar to residual connections!
- Replacing the arithmetic mean in Global Pooling is complex

upna

Summary of paper 3

- Combining the arithmetic mean with other pooling operators improves performance
  - Effect similar to residual connections!
- Replacing the arithmetic mean in Global Pooling is complex
  - A correct gradient flow is critical for the optimization.

upna

Summary of paper 3

- Combining the arithmetic mean with other pooling operators improves performance
  - Effect similar to residual connections!
- Replacing the arithmetic mean in Global Pooling is complex
  - A correct gradient flow is critical for the optimization.
- Exploiting chest x-ray data is a hard task.

Summary of paper 3

- Combining the arithmetic mean with other pooling operators improves performance
    - Effect similar to residual connections!
- Replacing the arithmetic mean in Global Pooling is complex
    - A correct gradient flow is critical for the optimization.
- Exploiting chest x-ray data is a hard task.
    - Benchmarking models against real-world datasets is important.

upna

Conclusion

Conclusion

- Monotonicity (increasingness) is important for feature fusion

Conclusion

- Monotonicity (increasingness) is important for feature fusion
    - We introduce a method to construct new increasing functions combining multiple operators.

Conclusion

- Monotonicity (increasingness) is important for feature fusion
    - We introduce a method to construct new increasing functions combining multiple operators.
- Preserving (or accentuating) high activations improves model performance:

upna

Conclusion

- Monotonicity (increasingness) is important for feature fusion
    - We introduce a method to construct new increasing functions combining multiple operators.
- Preserving (or accentuating) high activations improves model performance:
    - (a, b)-grouping functions can outperform classic operators.

Conclusion

- Monotonicity (increasingness) is important for feature fusion
    - We introduce a method to construct new increasing functions combining multiple operators.
- Preserving (or accentuating) high activations improves model performance:
    - (a, b)-grouping functions can outperform classic operators.
- Chest X-ray imaging prediction is a complex task to solve through CNN analysis:

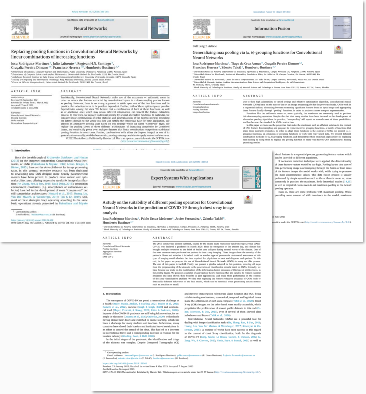
Conclusion

- Monotonicity (increasingness) is important for feature fusion
  - We introduce a method to construct new increasing functions combining multiple operators.
- Preserving (or accentuating) high activations improves model performance:
  - (a, b)-grouping functions can outperform classic operators.
- Chest X-ray imaging prediction is a complex task to solve through CNN analysis:
  - Below optimal results obtained with our modifications.

upna

Conclusion

PhD Project



Neural Networks 152 (2022) 380–393

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Replacing pooling functions in Convolutional Neural Networks by linear combinations of increasing functions

Iosu Rodriguez-Martinez [a], Julio Lafuente [a], Regivan H.N. Santiago [b], Graçaliz Pereira Dimuro [a,e], Francisco Herrera [c,d], Humberto Bustince [a,*]

ARTICLE INFO

ABSTRACT

Information Fusion 99 (2023) 101893

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Generalizing max pooling via (a, b)-grouping functions for Convolutional Neural Networks

Iosu Rodriguez-Martinez [a], Tiago da Cruz Asmus [b], Graçaliz Pereira Dimuro [a,c], Francisco Herrera [d], Zdenko Takáč [e], Humberto Bustince [a,*]

ARTICLE INFO

ABSTRACT

Expert Systems With Applications 235 (2024) 121162

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A study on the suitability of different pooling operators for Convolutional Neural Networks in the prediction of COVID-19 through chest x-ray image analysis

Iosu Rodriguez-Martinez [a], Pablo Ursua-Medrano [a], Javier Fernandez [a], Zdenko Takáč [b], Humberto Bustince [a,*]

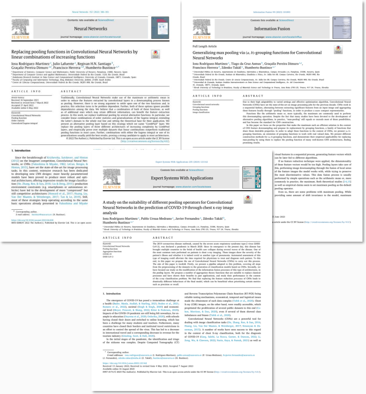
ARTICLE INFO

ABSTRACT

upna

Conclusion

PhD Project


Neural Networks
Replacing pooling functions in Convolutional Neural Networks by linear combinations of increasing functions

Information Fusion
Generalizing max pooling via (a, b)-grouping functions for Convolutional Neural Networks

Expert Systems With Applications
A study on the suitability of different pooling operators for Convolutional Neural Networks in the prediction of COVID-19 through chest x-ray image analysis

Collaborations


Information Fusion
A fusion method for multi-valued data

Fuzzy Sets and Systems
Affine construction methodology of aggregation functions

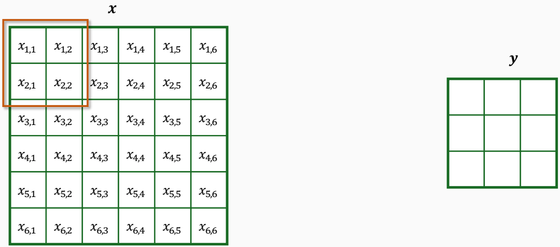
Information Sciences
A constructive framework to define fusion functions with floating domains in arbitrary closed real intervals

Fuzzy Sets and Systems
Extremal values-based aggregation functions

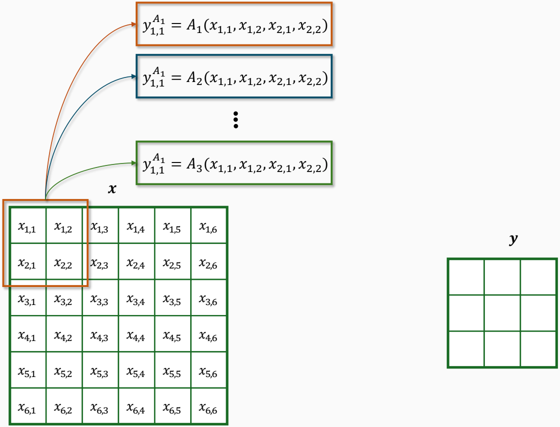
upna

Future research lines

Future lines (I)

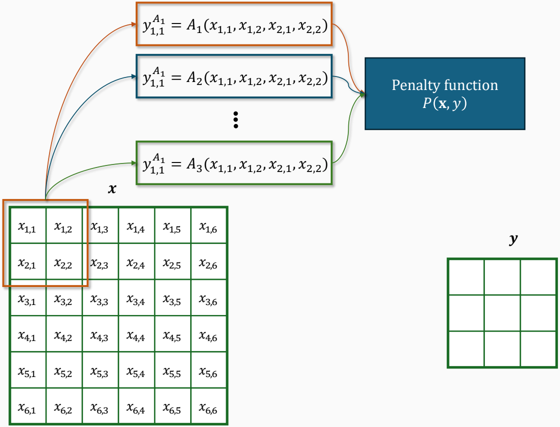


Future lines (I)



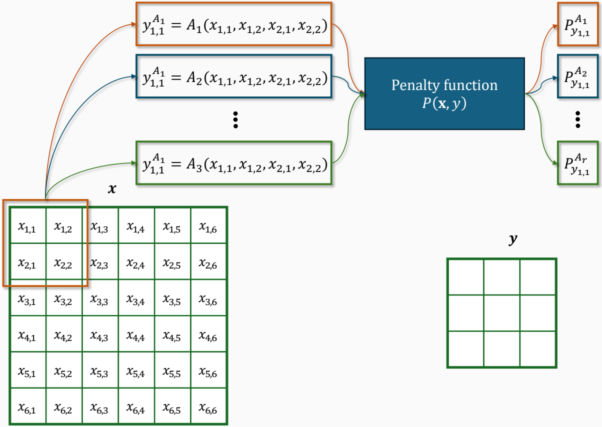
upna

Future lines (I)



Bustince, H., Beliakov, G., Dimuro, G. P., Bedregal, B., & Mesiar, R. (2017). On the definition of penalty functions in data aggregation. *Fuzzy Sets and Systems*, 323, 1-18.

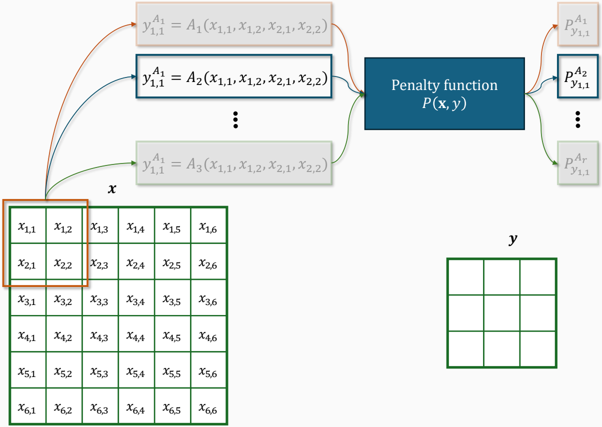
Future lines (I)



Bustince, H., Beliakov, G., Dimuro, G. P., Bedregal, B., & Mesiar, R. (2017). On the definition of penalty functions in data aggregation. *Fuzzy Sets and Systems*, 323, 1-18.

upna

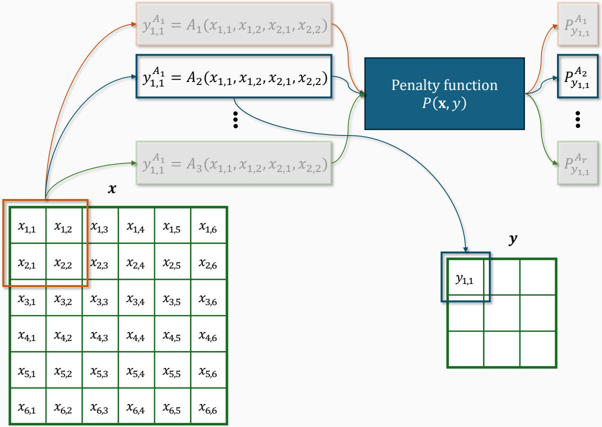
Future lines (I)



Bustince, H., Beliakov, G., Dimuro, G. P., Bedregal, B., & Mesiar, R. (2017). On the definition of penalty functions in data aggregation. *Fuzzy Sets and Systems*, 323, 1-18.

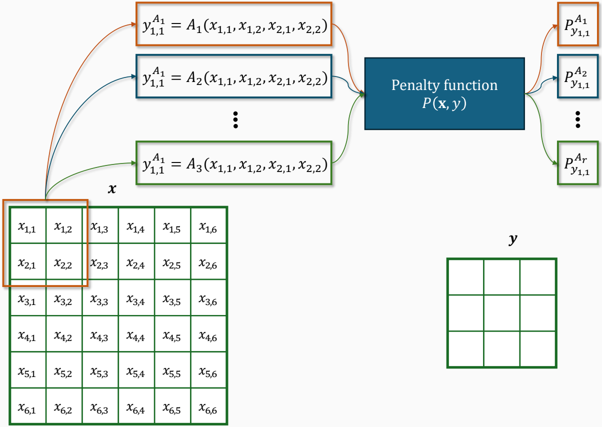
upna

Future lines (I)

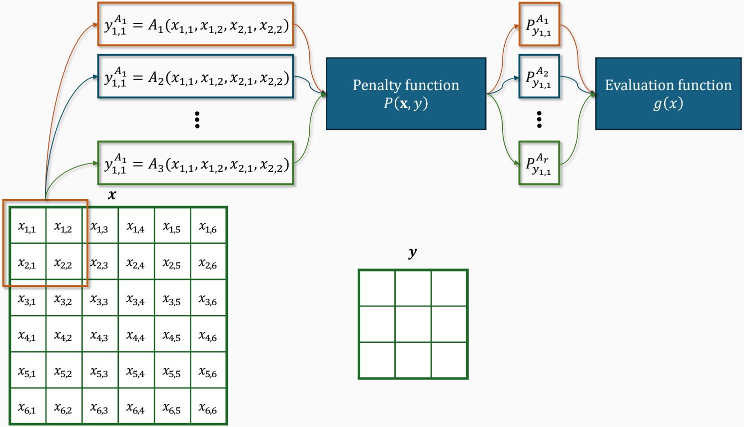


upna

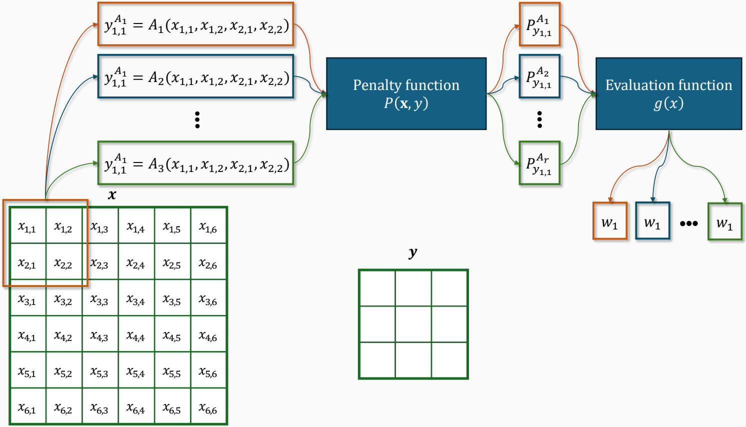
Future lines (I)



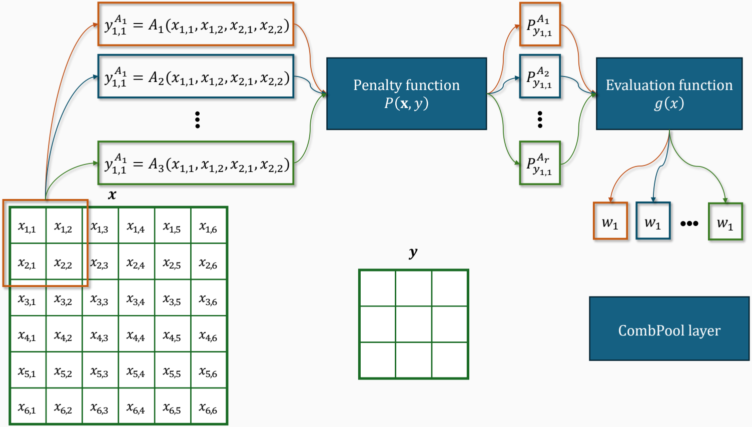
Future lines (I)



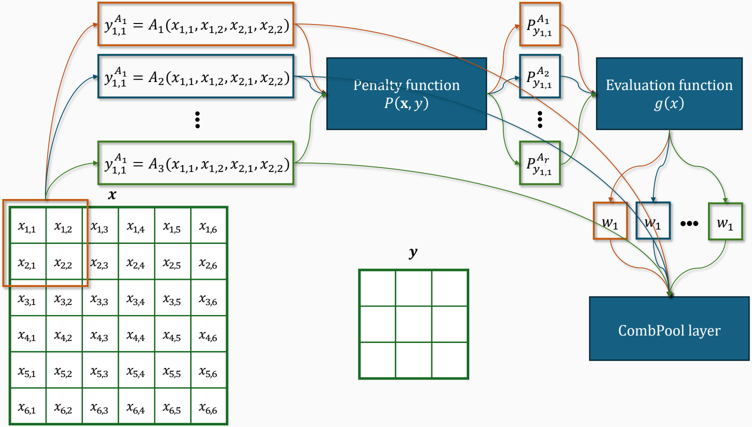
Future lines (I)



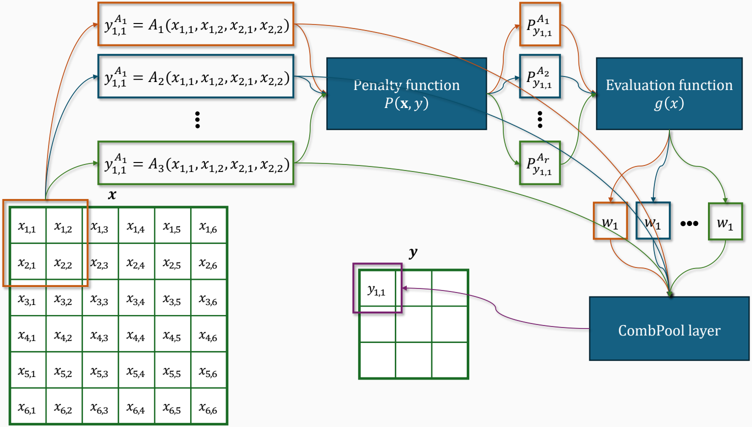
Future lines (I)



Future lines (I)



Future lines (I)



upna

Future lines (II)

- Improving upon Global Average pooling is non-trivial

[12] Papčo, M., Rodríguez-Martínez, I., Fumanal-Idocin, J., Altalhi, A. H., & Bustince, H. (2021). A fusion method for multi-valued data. *Information Fusion*, 71, 1-10.

[13] Halaš, R., Mesiar, R., Kolesárová, A., Saadati, R., Herrera, F., Rodriguez-Martinez, I., & Bustince, H. (2024). Extremal values-based aggregation functions. Fuzzy Sets and Systems, 493, 109097.

Future lines (II)

- Improving upon Global Average pooling is non-trivial
  - Could other averaging functions be used?

[12] Papčo, M., Rodríguez-Martínez, I., Fumanal-Idocin, J., Altalhi, A. H., & Bustince, H. (2021). A fusion method for multi-valued data. *Information Fusion*, 71, 1-10.

[13] Halaš, R., Mesiar, R., Kolesárová, A., Saadati, R., Herrera, F., Rodriguez-Martinez, I., & Bustince, H. (2024). Extremal values-based aggregation functions. Fuzzy Sets and Systems, 493, 109097.

Future lines (II)

- Improving upon Global Average pooling is non-trivial
  - Could other averaging functions be used?
  - Moderate-deviation functions have offered good results[12]

[12] Papčo, M., Rodríguez-Martínez, I., Fumanal-Idocin, J., Altalhi, A. H., & Bustince, H. (2021). A fusion method for multi-valued data. *Information Fusion*, 71, 1-10.

[13] Halaš, R., Mesiar, R., Kolesárová, A., Saadati, R., Herrera, F., Rodriguez-Martinez, I., & Bustince, H. (2024). Extremal values-based aggregation functions. Fuzzy Sets and Systems, 493, 109097.

Future lines (II)

- Improving upon Global Average pooling is non-trivial
  - Could other averaging functions be used?
  - Moderate-deviation functions have offered good results[12]
- Further exploit the importance of high activations

[12] Papčo, M., Rodríguez-Martínez, I., Fumanal-Idocin, J., Altalhi, A. H., & Bustince, H. (2021). A fusion method for multi-valued data. *Information Fusion*, 71, 1-10.

[13] Halaš, R., Mesiar, R., Kolesárová, A., Saadati, R., Herrera, F., Rodriguez-Martinez, I., & Bustince, H. (2024). Extremal values-based aggregation functions. Fuzzy Sets and Systems, 493, 109097.

Future lines (II)

- Improving upon Global Average pooling is non-trivial
    - Could other averaging functions be used?
    - Moderate-deviation functions have offered good results[12]
- Further exploit the importance of high activations
    - Define further aggregations based on extremal values[13].

[12] Papčo, M., Rodríguez-Martínez, I., Fumanal-Idocin, J., Altalhi, A. H., & Bustince, H. (2021). A fusion method for multi-valued data. *Information Fusion*, 71, 1-10.

[13] Halaš, R., Mesiar, R., Kolesárová, A., Saadati, R., Herrera, F., Rodriguez-Martinez, I., & Bustince, H. (2024). Extremal values-based aggregation functions. Fuzzy Sets and Systems, 493, 109097.

Future lines (II)

Fuzzy Sets and Systems 493-494 (2024) 109097

Contents lists available at ScienceDirect

Fuzzy Sets and Systems

journal homepage: www.elsevier.com/locate/fss

Extremal values-based aggregation functions

Radomír Halaš [a,*], Radko Mesiar [b,a], Anna Kolesárová [b], Reza Saadati [c], Francisco Herrera [d], Iosu Rodriguez-Martinez [e], Humberto Bustince [e,f]

[a] Department of Algebra and Geometry, Faculty of Science, Palacký University Olomouc, Czechia
[b] Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology, Radlinského 11, 810 05 Bratislava, Slovakia
[c] School of Mathematics, Iran University of Science and Technology, Narmak, Tehran, Iran
[d] Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, 18071, Granada, Spain
[e] Departamento de Estadística, Informática y Matemáticas, Universidad Pública de Navarra, Campus Arrosadia s/n, 31 006, Pamplona, Spain
[f] Insitute of Smart Cities, Universidad Pública de Navarra, Campus Arrosadia s/n, 31 006, Pamplona, Spain

A R T I C L E I N F O

Keywords:
Aggregation function
Choquet integral
Classification
Extended aggregation function
Ordered weighted quasi-arithmetic mean
Sugeno integral
t-conorm
t-norm
Symmetric discrete fuzzy measure

A B S T R A C T

We introduce and study aggregation functions based on extremal values, namely extended (l,u)-aggregation functions whose outputs only depend on a fixed number l of extremal lower input values and a fixed number u of extremal upper input values, independently of the arity of the input n-tuples ($n \geq l + u$). We discuss several general properties of (l,u)-aggregation functions and we study special (l,u)-aggregation functions with neutral element, including t-conorms, t-norms and uninorms. We also study (l,u)-aggregation functions defined by means of integrals with respect to discrete fuzzy measures, as well as (l,u)-ordered weighted quasi-arithmetic means based on appropriate weighting vectors. We also stress some generalizations based on recently introduced new types of monotonicity. Some possible applications are sketched, too.

1. Introduction[1]

In aggregation processes we often meet situations where some extremal input values are either excluded from a sample of values to be aggregated or oppositely, only some extremal values are aggregated. The first type of aggregation is applied, for example, in some sport events where the obtained maximal and minimal evaluations are not considered. As a typical example we recall the way of evaluation of the jumping style of competitors in ski-jumping. The score of each of five judges evaluating their style can reach a maximum of twenty points, but the lowest and the highest style scores are disregarded and only the sum of the remaining three scores is considered for the final evaluation. Similarly, in some sports considering several competitions for obtaining the global evaluation, such as yachting, often only some extremal (maximal/minimal) results are considered. In aggregation theory such an approach is usually modeled by particular OWA operators [36]. In the latter case, we can exemplify aggregation by means of some t-conorms and related functions [1,17] or by particular integrals [28,34]. For example, let us mention aggregation performed by extremal t-conorms: in the case when the smallest t-conorm $S_M = \max$ is applied, the result depends on the maximal input value only, and when a sample is aggregated by the greatest t-conorm S_D (the drastic sum), the result only depends on two maximal input values. Also, in statistics, the mid-range parameter (measuring the central tendency of a sample) which is defined for any statistical sample $\mathbf{x} = (x_1, \ldots, x_n)$

* Corresponding author.
E-mail addresses: radomir.halas@upol.cz (R. Halaš), radko.mesiar@stuba.sk (R. Mesiar), anna.kolesarova@stuba.sk (A. Kolesárová), rsaadati@eml.cc (R. Saadati), herrera@ugr.es (F. Herrera), iosu.rodriguez@unavarra.es (I. Rodriguez-Martinez), bustince@unavarra.es (H. Bustince).

[1] The paper is an extended version of the original talk presented at IFSA 2021 conference, see [21].

https://doi.org/10.1016/j.fss.2024.109097
Received 8 March 2024; Received in revised form 26 June 2024; Accepted 25 July 2024

Available online 30 July 2024

Future lines (II)

- Improving upon Global Average pooling is non-trivial
    - Could other averaging functions be used?
    - Moderate-deviation functions have offered good results[14]
- Further exploit the importance of high activations
    - Define further aggregations based on extremal values[15].
- Replace other feature fusion processes

[14] Papčo, M., Rodríguez-Martínez, I., Fumanal-Idocin, J., Altalhi, A. H., & Bustince, H. (2021). A fusion method for multi-valued data. *Information Fusion*, 71, 1-10.

[15] Halaš, R., Mesiar, R., Kolesárová, A., Saadati, R., Herrera, F., Rodriguez-Martinez, I., & Bustince, H. (2024). Extremal values-based aggregation functions. Fuzzy Sets and Systems, 493, 109097.

upna

Thanks for your attention

Time for questions

iosu.rodriguez@unavarra.es
Departament of Statistics, Computer Science and Mathematics (UPNA)
Grupo de Inteligencia Artificial y Razonamiento Aproximado (GIARA)

Acknowledgements I. Rodriguez-Martinez and H. Bustince are supported by the grant PID2022-136627NB-I00. Z. Takáč is supported by grant VEGA 1/0267/21. F. Herrera is supported by the grant P18-FR4961. I. Rodriguez-Martinez is also financially supported by Tracasa Instrumental (iTracasa) and Gobierno de Navarra - Departamento de Universidad, Innovación y Transformación Digital.