HERITAGE

Image recognition of architectural heritage images

Alfredo Galli & Riccardo Rubini

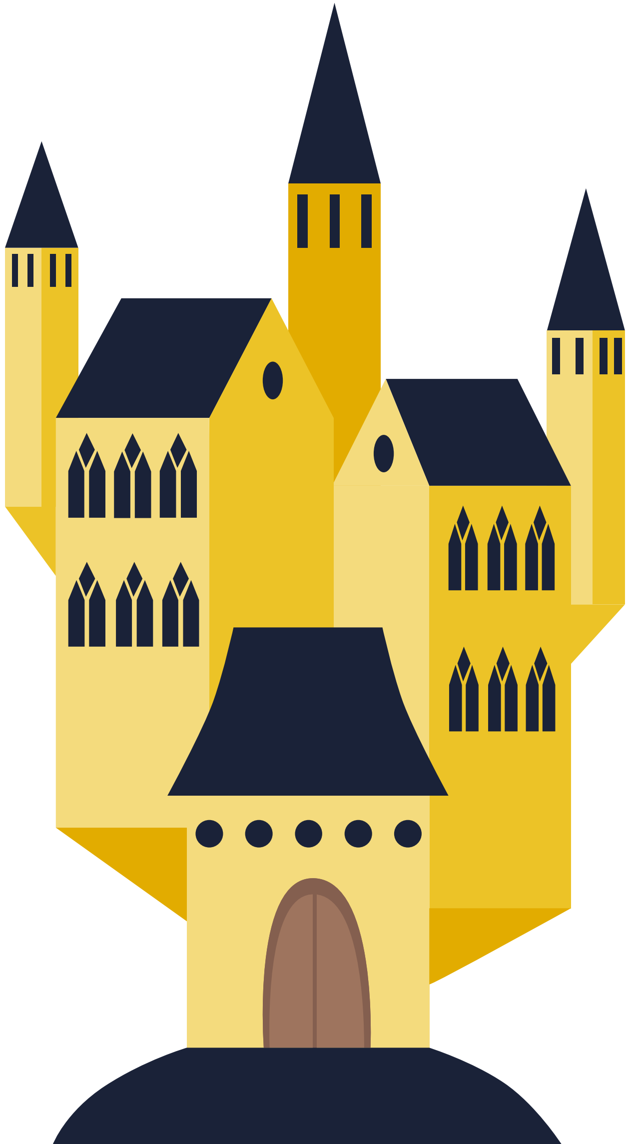
Exam Project for Foundations of Deep Learning
A.A. 2021/2022

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Dataset description

Dataset consists of **10'235** training images (128x128) classified in 10 different categories:

Altar 829 img.	**Apse** 514 img.	**Bell Tower** 1059 img.	**Column** 1919 img.	**Dome out.** 1177 img.
Dome in. 616 img.	**Buttress** 407 img.	**Gargoyle** 1571 img.	**Stained Glass** 1033 img.	**Vault** 1110 img.



Images example

Altar

Altar

Apse

Apse

Column

Bell Tower

Bell Tower

Dome In.

Dome Out.

Buttress

Gargoyle

Gargoyle

Stained Glass

Vault

Vault

Data Preprocessing

During the training process, the 10'235 train images have been splitted in:

- **Train Set:** 8'188 images – **80%**
- **Validation:** 2'047 images – **20%**

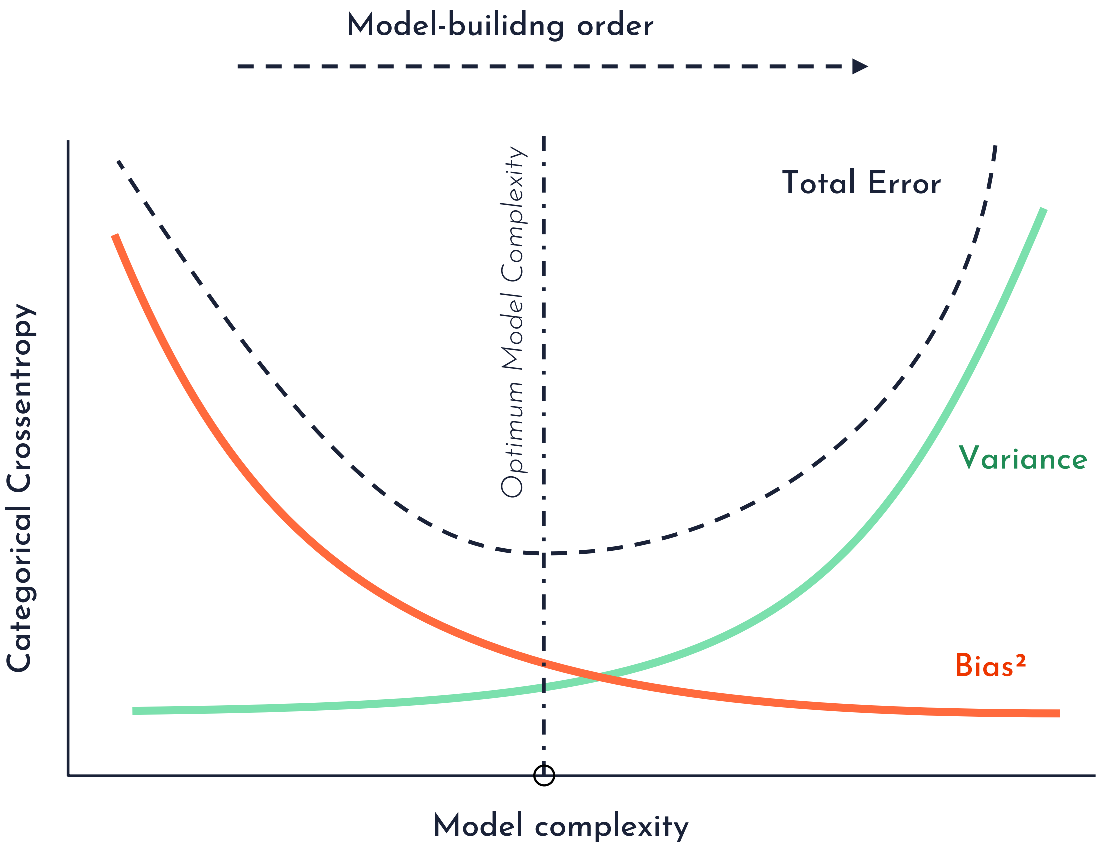
In the final step **1'404 images** in the **test set** have been used in order to evaluate in a robust way the model performance.

Following the best practices, **the data have been normalized** in a [0,1] range. After thate, we have created **batch of 32 images**.

Model-building strategies

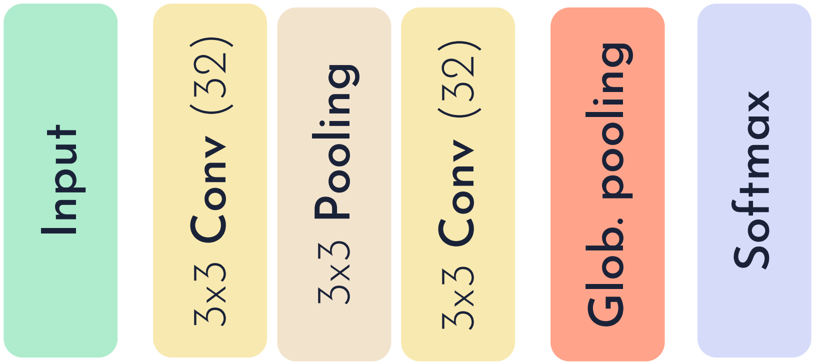
We have started **creating shallow neural network** and then we have made them **deeper towards the end** of the model-building process.

The results have been evaluated on the validation set. An **early stop loss criterion** was used on it (with patience = 4) in order to reduce the risk of overfitting and decrease the computational cost of the training phase.



Shallowest Model

20'042 parameters



Performance

	Loss	Accuracy
Train	0,678	77,7%
Validation	0,812	71,1%

Deepest Model

1'927'019 parameters

Performance

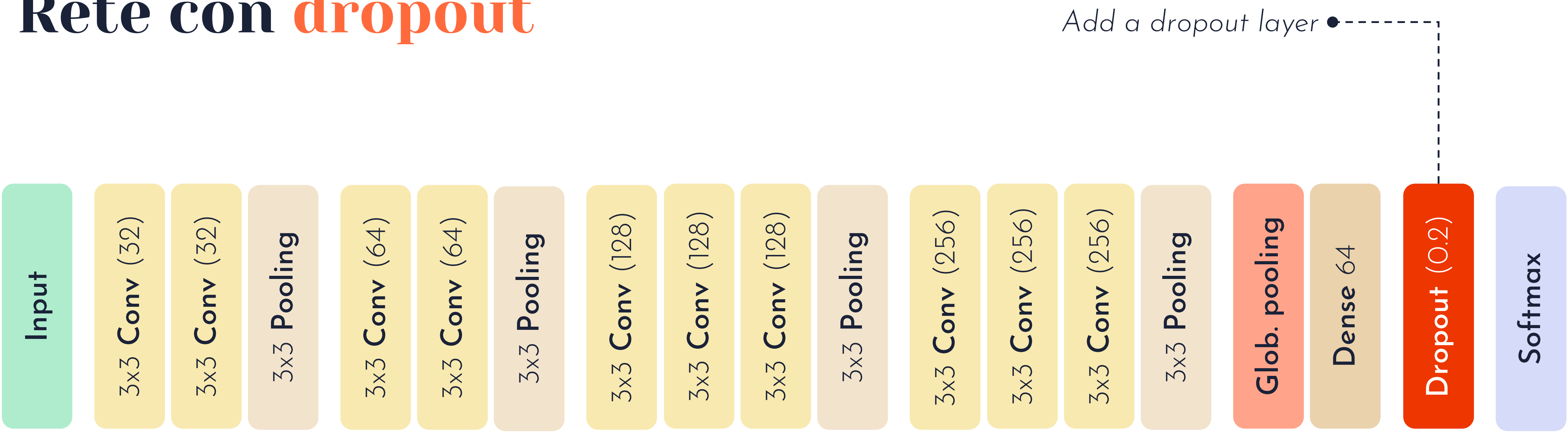
	Loss	Accuracy
Train	0,429	85,7%
Validation	0,636	77,9%

Neural Network
Regularization

The increase of the parameters number has **gradually led to a general improvement** in the performance of the model. However, in the deepest models we have arrived **near to an overfitting problem**.

In order to **reduce overfitting** and help the model to **generalize well**, we have tried some of the most common **regularization techniques**.

Rete con dropout

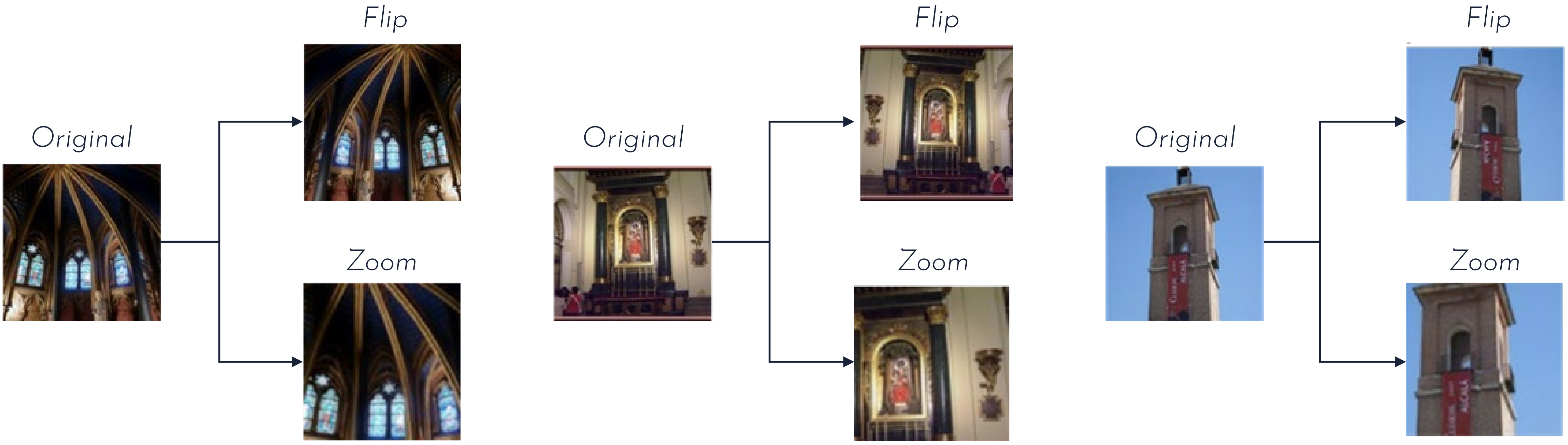


Performance

	Loss	Accuracy
Train	0,281	90,6%
Validation	0,561	82,8%

Data augmentation

We have tried another regularization techniques on the most complex models, the data augmentation, that help also to **increase the training images number**. In our work, we have used two transformation: **oriziontal flip and random zoom**, thus **tripling the training dataset**.



Rete con data augmentation

1'091'242 parametri

Performance

	Loss	Accuracy
Train	0,338	88,6%
Validation	0,415	86,1%

Si è scelta una rete meno complessa rispetto alle precedenti perché l'aumento del numero dei parametri è parso non portare grandi benefici in termini di diminuzione della loss (accuracy sempre introno all'83-86%)

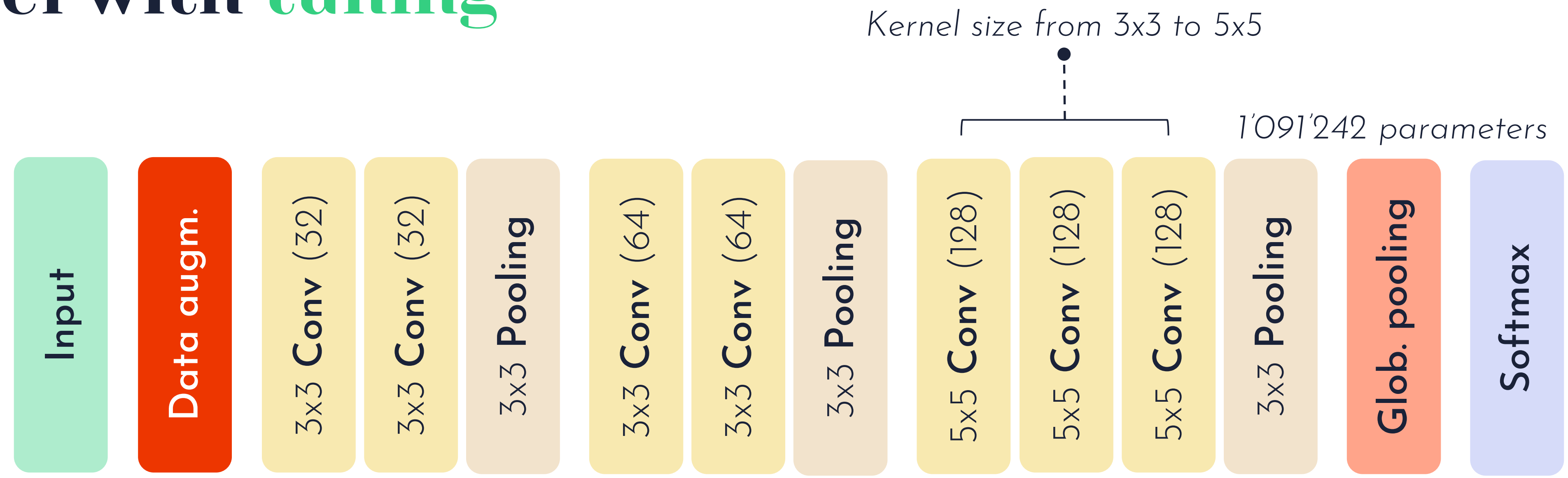
Hyperparameter Optimization

Once the network architecture and the best regularization method were defined, we moved on to the hyperparameter tuning phase. In our case we tried to find the best combination between:

1. Number of ***convolutional filters***
2. The ***kernel dimension*** of *the last convolutional part*
3. Type of ***padding***
4. Value of ***learning rate***

We have tried different combinations by a **random search** approach.

Model with tuning



Performance

	Loss	Accuracy
Train	0,184	94,1%
Validation	0,416	89,8%

Confusion Matrix on test set

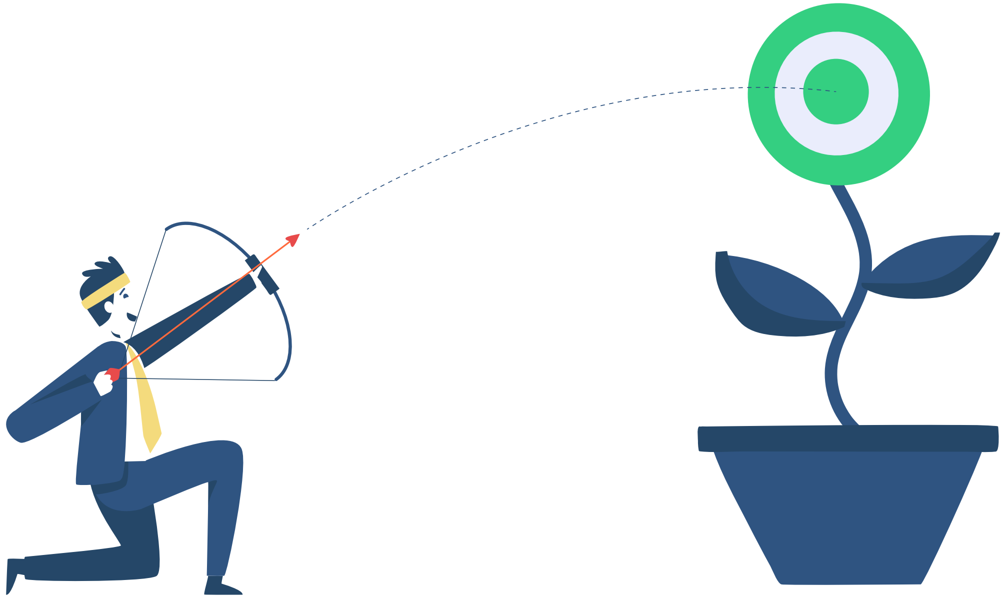
PREDICTED CATEGORY

ORIGINAL CATEGORY	Altar	Apse	Bell Tower	Column	Dome In.	Dome out.	Buttress	Gargoyle	St.Glass	Vault
Altar	89	8	2	12	6	1	0	9	1	12
Apse	0	38	1	0	1	7	2	1	0	0
Bell Tower	0	6	115	8	0	28	2	11	0	0
Column	0	17	3	151	6	18	1	13	0	1
Dome in.	0	0	0	1	60	0	0	5	0	3
Dome out.	0	3	2	0	0	132	0	5	0	0
Buttress	1	5	2	3	1	0	40	18	0	0
Gargoyle	0	3	2	4	8	0	0	221	0	2
St. Glass	3	2	2	1	4	0	0	2	133	3
Vault	6	1	0	1	4	1	0	6	1	143

Test set metrics

accuracy

79,9%



STRATIFIED METRICS

CATEGORY	*Precision*	*Recall*	*F1*
Altar	0,90	0,64	0,74
Apse	0,46	0,76	0,57
Bell Tower	0,89	0,68	0,77
Column	0,83	0,72	0,77
Dome in.	0,67	0,87	0,75
Dome out.	0,71	0,93	0,80
Buttress	0,89	0,57	0,70
Gargoyle	0,76	0,92	0,83
St. Glass	0,99	0,89	0,93
Vault	0,87	0,88	0,87

Conclusion and future work

General performance

The model is able to correctly classify approximately 8 test images out of 10

Performance by category

9 out of 10 categories have a good F1 score, above 0.70. The only problematic category is apse, for which the score is 0.57

Data Augmentation

Given the positive contribution, an expansion of the data augmentation could be evaluated, both for generated images and for type of transformations

Transfer Learning

Since the network used was created from scratch, the use of transfer learning could be studied by exploiting some of the most famous networks in the literature