Équipes Traitement
de l'Information
et Systèmes

ENSEA
Ecole Nationale
Supérieure
de l'Electronique
et de ses Applications

Recognition of age and gender by deep learning

In-depth analysis of the convolutional neural network

Author: Cristian Steeven Maquilon Mejia

Supervisor: Michel Chapron

August 28, 2023

ACKNOWLEDGEMENTS

Firstly, thanks to the research professor and tutor of this work, Michel Chapron, for the knowledge provided, to the director for his management Olivier Romain also to the ETIS research group for the trust granted, to ENSEA, which provides its facilities to contribute to the investigations carried out by the research group. And last and most importantly, to all my family who, even from the distance, always contribute to my successes and support me in defeats.

CONTENTS

1 Introduction

The detection and prediction of facial features, such as age and gender, are fundamental tasks in computer vision and image processing. These applications have a wide range of practical uses, from age identification for access control to the personalization of online content and advertising. The Convolutional Neural Network (CNN) has emerged as a powerful tool in these tasks, enabling the extraction of features from images with exceptional accuracy.

Our research covers three CNN models, each with a different configuration: a model with 128x128 pixel input images, another with 192x192 pixel images, and a third model that incorporates database modifications. In addition to training and evaluating these models, our main objective is to analyze the relationships between the filters in the different layers of the convolutional neural networks. These analyzes were carried out with the purpose of obtaining relevant information about the similarities of the filters obtained during training by CNN.

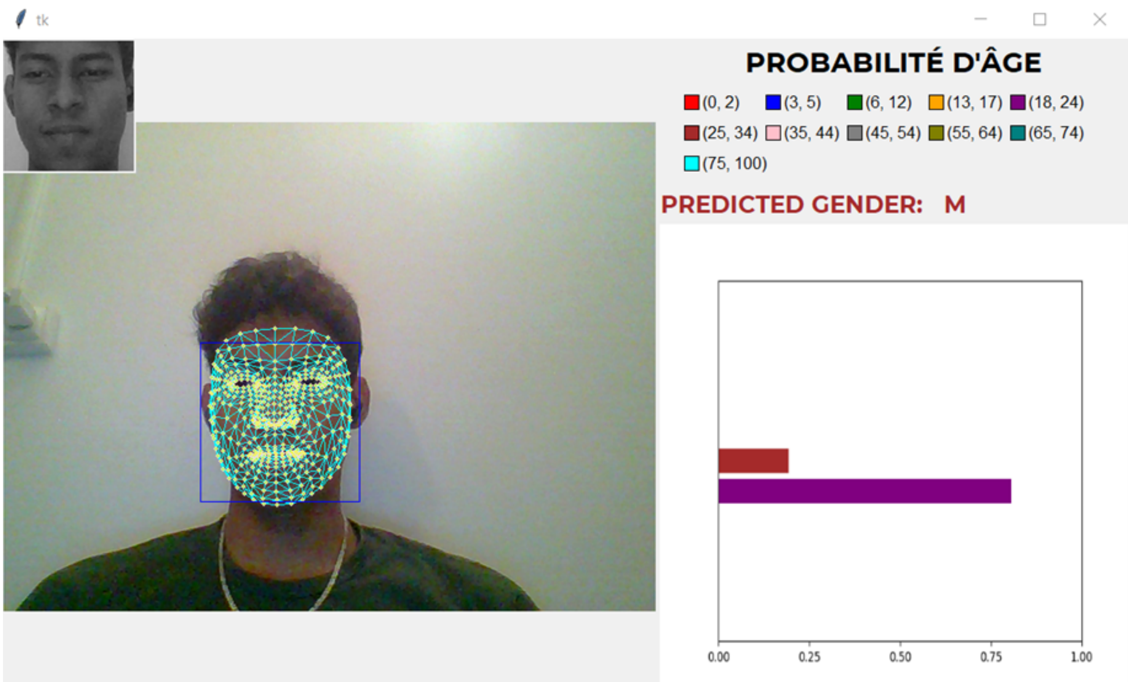


Figure 1.1. small functional application of age and gender prediction

The first model was applied to a small python application that predicted age and gender in real time helped by computer vision, however, it could be improved. As a result of the above, this project emerged as a solution to the failures of the old model to optimize and improve the results.

This report represents a significant step in the understanding and application of convolutional neural networks in age- and gender-related computer vision tasks, and is expected to contribute to the continued growth of this research area, especially in the search for improvements. in learning, training time and depth of learning of CNNs.

The central focus of our study is to identify potential redundancies in order to optimize learning processes, reduce learning time, and improve the learning depth of CNNs. It is important since the search for strategies to optimize these dimensions is essential in the construction of effective and efficient computer vision models.

The report is organized as follows: A section dedicated to Etis, its work groups and study areas, "Rapport d'étonnement" where my time at Etis is detailed, the Methodology section, we detail the construction of the models, the training and analysis techniques used. Subsequently, in the Results and Discussion section, we present the results obtained, analyze the relationships between the filters, and discuss the implications of our findings. Finally, we conclude by highlighting the contributions of our study and possible future directions in the application of CNN in the prediction of age and gender from facial images, with a focus on the optimization of learning processes.

2 ETIS

ETIS is a joint research department between CYU Cergy Paris University, ENSEA Graduate School of Electrical Engineering and CNRS/INS2I. The department is currently headed by Prof. Olivier Romain assisted by Prof. Lola Canamero.

ETIS develops research in the field of the theory of information with both theoretical and experimental activities in order to allow information processing systems to acquire capacities of autonomy. Autonomy is considered both in terms of learning and adaptation to the environment (including users) as well as making decision that includes low energy consumption and computing power for example.

ETIS designed systems perform intelligent processing which is adaptable to increasing complexity. The concerned areas are reconfigurable chip systems, data analysis, image indexing, developmental robotics, information theory and telecommunications. Learning and adaptation algorithms based on data constitue the core of the developed systems.

The ETIS laboratory is at the heart of the current AI revolutionLearning and adaptation algorithms based on data constitue the core of the developed systems.

2.1 CELL

CELL is a multidisciplinary team working in the field of Electronics, Signal and Image Processing for the Modeling and Design of Reconfigurable, Communicating, Reliable and Intelligent Embedded Systems.

A particular emphasis is put on the design and use of Machine Learning techniques in contexts specific to data analysis and data processing problems, with a focus on the explainability and transparency of such algorithms.

Applications include social media content and graph analytics, semantic web data integration and analysis, pattern detection in image, video and multimodal data, search engines for different types of data, knowledge extraction from data, data privacy, etc.

The application domains are vast, such as cultural heritage, security, opinion mining, health, etc.

CELL research is now divided into three scientific axes representing the two levels of activity of the team, namely an upstream technological level and an application level linked to the socio-economic fabric. These are addressing problems with societal stakes in line with the areas of expertise and values of the team.

Research areas:

- Smart Embedded Systems for Health.
- Non-Conventional Sensors.
- The Wizarde Project.
- Reliability and ECC
- Circuits and RF Systems

2.2 NEURO

Our group pushes forward embodiment and bio-inspiration toward cognitive and social robots, autonomous and embodied AI. We investigate biological models for human-level cognition and interact with developmental psychologists and neuroscientists.

Research areas:

- BRAIN.
- COGPERCEPTION.
- AUTONOMOUS.
- HEALTH.
- BODY.

2.3 MIDI

The MIDI group works in the field of Big Data Management and Analytics, addressing topics that include integrating, mining, analyzing and querying large volumes of various types of data, ranging from structured and semi-structured data, to spatiotemporal data, text, image, video and 3D models.

A particular emphasis is put on the design and use of Machine Learning techniques in contexts specific to data analysis and data processing problems, with a focus on the explainability and transparency of such algorithms.

Applications include social media content and graph analytics, semantic web data integration and analysis, pattern detection in image, video and multimodal data, search engines for different types of data, knowledge extraction from data, data privacy, etc.

The application domains are vast, such as cultural heritage, security, opinion mining, health, etc.

Research areas:

- Responsible Data Science.
- Data Integration and Analytics for multiple modalities.
- Distributed, Online and Deep Learning.

2.4 ICI

The ICI group's research focus is on wireless communications, information theory, signal processing and imaging.

The key application areas covered by the group's research output lie primarily on 5G/6G, IoT and machine learning for communications; coding and information theoretic security; networking and edge computing; imaging and modelization.

Research areas

The ICI group's research focus is on core topics on wireless communications, information theory, signal processing and MAC / network slicing, emerging imaging and modelling.

Current research topics are pertinent to IoT verticals (with emphasis on V2X and Smart factories) and lie primarily on the following domains:

1. **Wireless communications, B5G and 6G / PHY aspects**
2. **Coding and information theory**
3. **Networking slicing and caching, including**
4. **Security for wireless communications including**
5. **Imaging and modelling**

2.5 Rapport d’étonnement

As an exchange student, my process with ETIS was very favourable, the recruitment process was very simple and not at all complex, as was the initiation period for the internship.

Since my arrival at Etis, I have deepened my knowledge about convolutional neural networks, image processing, programming work in python and data science, knowledge that I consider important for my development as a future engineer because they are very established jobs in daily life.

Working at etis was an interesting experience being located in the ENSEA offices, mobility was not a problem, my tutor was always aware of my progress.

In general, I consider that this opportunity was rewarding and I would recommend it to people who really like research and are willing to work for the development of a possible discovery.

3 METHODS

3.1 Construction of the Convolutional Neural Network (CNN)

To perform the task of predicting a person's age and gender, a Convolutional Neural Network (CNN) was implemented. CNNs are a type of deep neural network architecture specifically designed for processing image data. This section will provide an overview of how the CNN was constructed and trained for this project, using the UTKFace database.

3.1.1 Data Acquisition and Preprocessing

The first step in building the CNN was acquiring the UTKFace database, which contains labeled facial images along with information about the age and gender of individuals. This database is widely recognized and used in the computer vision community for tasks related to age and gender.

Preprocessing tasks were carried out on this database, which included image normalization and resizing, to ensure that the input data was in optimal conditions for network training. The distribution of the database in terms of age is as follows:

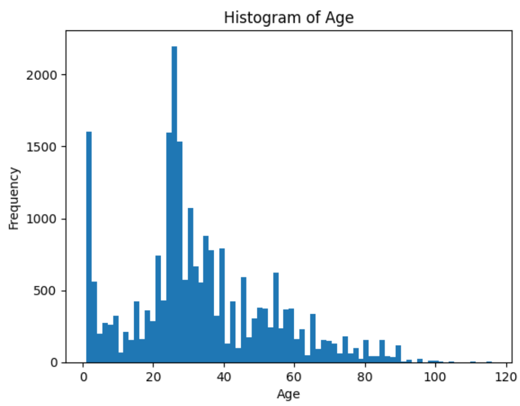


Figure 3.1. Age distribution from UTKFace database

3.1.2 CNN architecture

Three different CNN models were trained, each with a slightly different configuration:

Model 1 - 128x128

The first model was designed with an input image size of 128x128 pixels. This resolution was chosen as a starting point for initial exploration because this type of neural network is suitable for smaller input data.

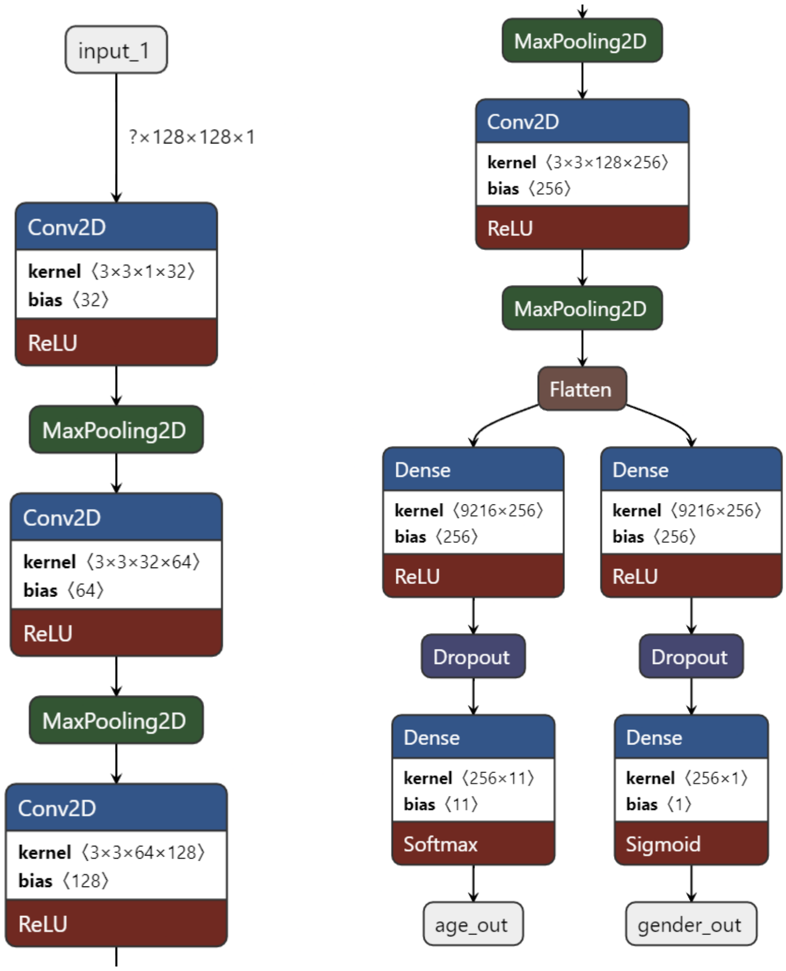


Figure 3.2. Neural structure of convolutional network Model 1 and 1.1

prediction:

```
1/1 [==============================] - 1s 742ms/step
Original Gender: 0 Original Age Range: (45, 54)
Predicted Gender: 0
Probabilities for each age range:
Age Range (0, 2): 0.00%
Age Range (3, 5): 0.00%
Age Range (6, 12): 0.00%
Age Range (13, 17): 0.00%
Age Range (18, 24): 0.00%
Age Range (25, 34): 1.29%
Age Range (35, 44): 0.24%
Age Range (45, 54): 98.48%
Age Range (55, 64): 0.00%
Age Range (65, 74): 0.00%
Age Range (75, 100): 0.00%
```

Figure 3.3. Moldel 1 prediction of the convolutional network

Model 1.1 - 128x128 (database modification)

In the second model, modifications were made to the UTKFace database based on a new concept suggested by the researcher and professor Chapron. This concept involved altering the least significant bits of the images. This modification reduced the training time by approximately 30% compared to the first model. Although the predictions vary slightly compared to the first model, the model remains reliable. Since it's the same model, the structure is identical to that shown in Figure 3.2.

Prediction:

```
1/1 [==============================] - 0s 147ms/step
Original Gender: 0 Original Age Range: (45, 54)
Predicted Gender: 0
Probabilities for each age range:
Age Range (0, 2): 0.00%
Age Range (3, 5): 0.00%
Age Range (6, 12): 0.00%
Age Range (13, 17): 0.00%
Age Range (18, 24): 0.00%
Age Range (25, 34): 35.41%
Age Range (35, 44): 33.39%
Age Range (45, 54): 31.20%
Age Range (55, 64): 0.00%
Age Range (65, 74): 0.00%
Age Range (75, 100): 0.00%
```

Figure 3.4. Moldel 1.1 prediction of the convolutional network

Modelo 2 - 192x192

The third model was configured to accept higher-resolution input images with dimensions of 192x192 pixels. This variation allowed for an evaluation of whether a higher resolution improved the network's performance.

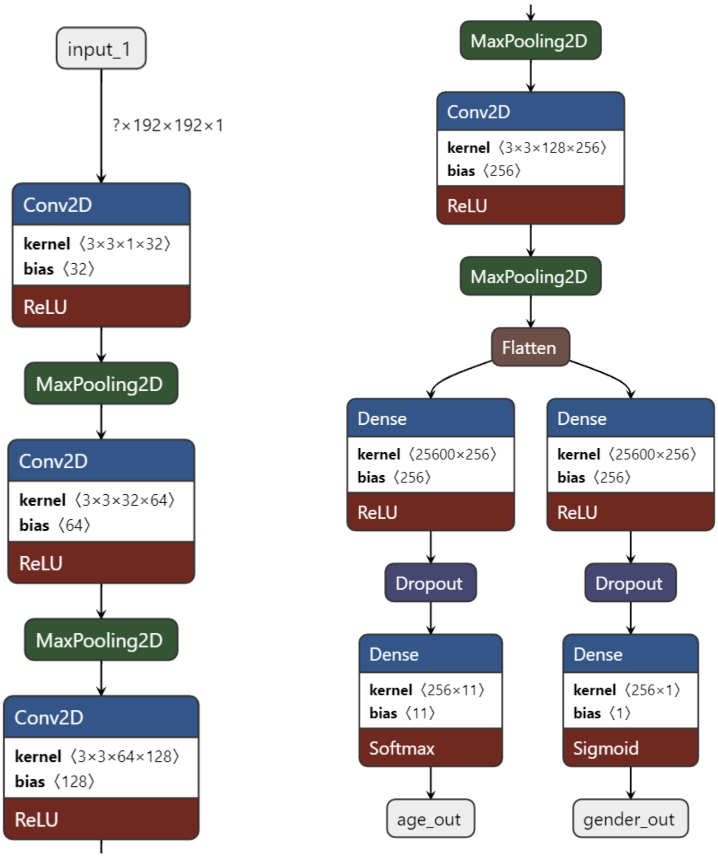


Figure 3.5. Neural structure of convolutional network

Prediction:

```
1/1 [==============================] - 2s 2s/step
Original Gender: 0 Original Age Range: (45, 54)
Predicted Gender: 0
Probabilities for each age range:
Age Range (0, 2): 0.00%
Age Range (3, 5): 0.00%
Age Range (6, 12): 0.00%
Age Range (13, 17): 0.00%
Age Range (18, 24): 0.00%
Age Range (25, 34): 79.55%
Age Range (35, 44): 20.45%
Age Range (45, 54): 0.00%
Age Range (55, 64): 0.00%
Age Range (65, 74): 0.00%
Age Range (75, 100): 0.00%
```

Figure 3.6. Moldel 2 prediction of the convolutional network

3.1.3 CNN Training

Each of the three models was trained using the UTKFace dataset modified according to the resolutions or modifications applied. Specific optimization techniques were employed to adjust the network's weights and minimize the loss function, using a combination of binary cross-entropy and categorical cross-entropy. This combination is appropriate for gender classification and age regression tasks. Additionally, the 'Adam' optimizer was used.

Multiple training epochs were conducted to allow each CNN to learn relevant patterns in the data and improve its predictive capacity. The process was closely monitored by tracking performance metrics on the validation set and detecting potential issues such as overfitting.

3.1.4 Performance Evaluation

Once the three CNN models were adequately trained, their performance was evaluated using the UTKFace test dataset. Additionally, specific metrics were calculated to assess the accuracy of age and gender prediction in each of the models.

Model Results

Below, the results obtained for the three trained models are summarized:

Model 1 - 128x128:

- Age prediction accuracy: 92.72%.
- Gender prediction accuracy: 100.0%.
- Training time: Approximately 47 minutes.

Model 1.1 - Modified Database:

- Age prediction accuracy: 69.79%.
- Gender prediction accuracy: 90.0%
- Training time: Approximately 31 minutes.

Model 2 - 192x192:

- Age prediction accuracy: 17.12%.
- Gender prediction accuracy: 40%.

- Training time: Approximately 1 hour and 20 minutes.

The accuracy measurement was conducted by making multiple predictions and calculating the average correctness for the various options.

3.2 Analysis methods

In this section, we will explain the analysis methods used to evaluate and understand the results of the Convolutional Neural Network (CNN) models trained in the previous section, section 3.1.

3.2.1 Correlation Matrix

The correlation matrix was used to measure the linear relationship between filters in different layers of the CNN. By calculating the correlation between filters, it was possible to identify if there were significant relationships between them, which could indicate redundancy or complementarity in the information they captured.

In this report, the correlation matrices provide detailed information about the relationships between filters in the first layer of each model. Each value in the matrix represents the correlation between two filters. A correlation close to 1 indicates a strong positive correlation, while a correlation close to -1 indicates a strong negative correlation. Analyzing these matrices helps us identify patterns and relationships among filters in these specific layers.

The correlation between two data sets X and Y is defined as:

$$\rho(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y}$$

Where:

- X and Y are data sets.
- $\rho(X,Y)$ represents the correlation between X and Y.
- n is the number of observations in the data sets.
- S_X is the standard deviation of X.
- S_Y is the standard deviation of Y.
- $\bar{X}$ is the mean of X.

- $\bar{Y}$ is the mean of Y.
- X_i and Y_i are the individual values in the data sets.

3.2.2 Infinity Norm Matrix

The Infinity Norm matrix was applied to assess the magnitude of the differences between the CNN filters. the information we receive from the infinity norm is important, as a high value of Infinity Norm could indicate significant divergence in the information being processed by the CNN layers or, conversely, redundancy.

The infinite norm of a vector X is defined as:

$$||X||_\infty = \max |X_i|$$

Where:

- X is a vector in space.
- $||X||_\infty$ represents the infinite norm of X.
- $\max_i |X_i|$ means taking the absolute maximum value of all elements X_i of vector X.

3.2.3 Cosine Similarity Matrix

Cosine similarity was used to measure the directional similarity between the filters in the various layers of the CNN. A high cosine similarity between two filters would indicate that their directions were similar, which could suggest that they were capturing similar features in the input images.

Cosine similarity produces a value that varies between -1 and 1. The closer the value is to 1, the more similar the two vectors are. A value of 1 means that the vectors are identical, while a value of -1 means that they are completely opposite. A value of 0 implies that the vectors are orthogonal or have no similarity.

The similarity of cosines of two vectors X and Y is defined as:

$$\text{Cosine Similarity}(X, Y) = \frac{X \cdot Y}{||X|| ||Y||}$$

Where:

- X and Y are vectors in space.

- $||X||$ represents the norm (length) of the vector X.
- $||Y||$ represents the norm (length) of the vector Y.
- $X \cdot Y$ represents the dot product between the vectors X and Y.

These analysis methods provided a deeper understanding of how the filters in the different layers of the CNN interact and are related, which contributed to a better understanding of how the neural network processed the information from the input images.

4 RESULTS AND DISCUSSION

4.1 Further Analysis: Applying Filters to Random Images

Visualization of the first layer kernels is essential to understand how Convolutional Neural Network (CNN) models are capturing key features of images. Additionally, to gain a deeper understanding of how the filters in our Convolutional Neural Network (CNN) models affect images, we conducted additional analysis. We applied these filters to a random image selected from the data set to see how they behaved and to identify noticeable patterns.

Modelo 1 - 128x128

Figure 4.1. Kernels and filtered images

Model 1.1 - 128x128 (database modification)

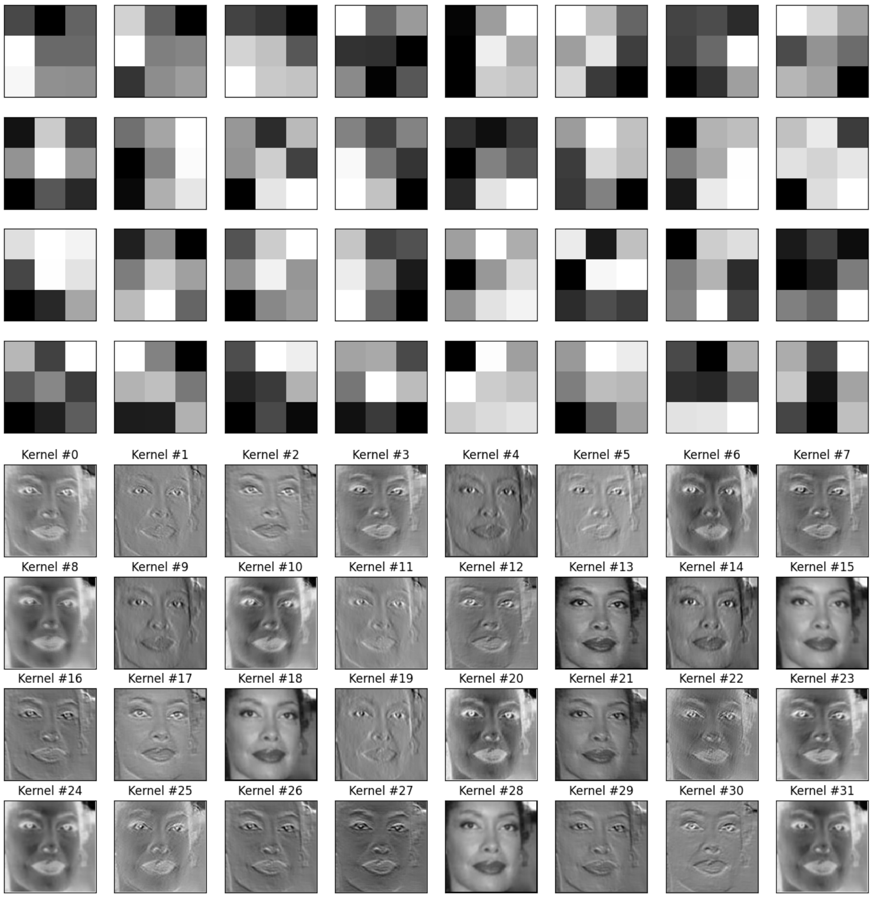


Figure 4.2. Kernels and filtered images

Modelo 2 - 192x192

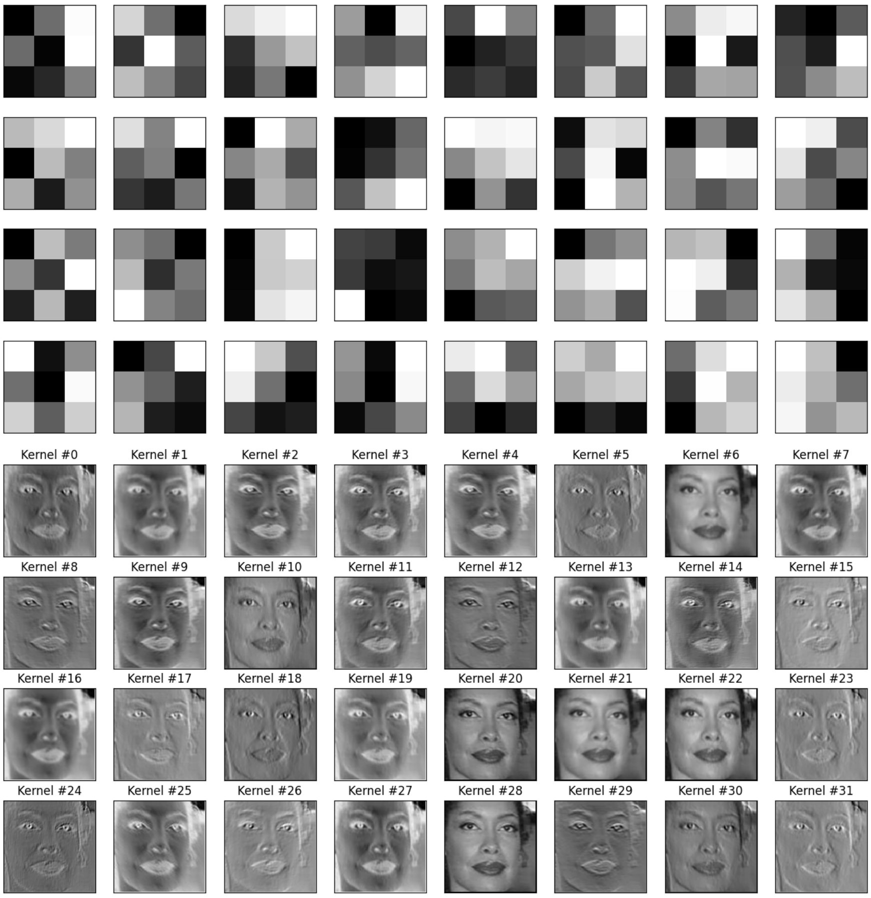


Figure 4.3. Kernels and filtered images

In one finding, we discovered that by inverting the channels of some filtered images, we got an image that was the same as another previously filtered image. This indicates a complex relationship between the filters and the information in the images.

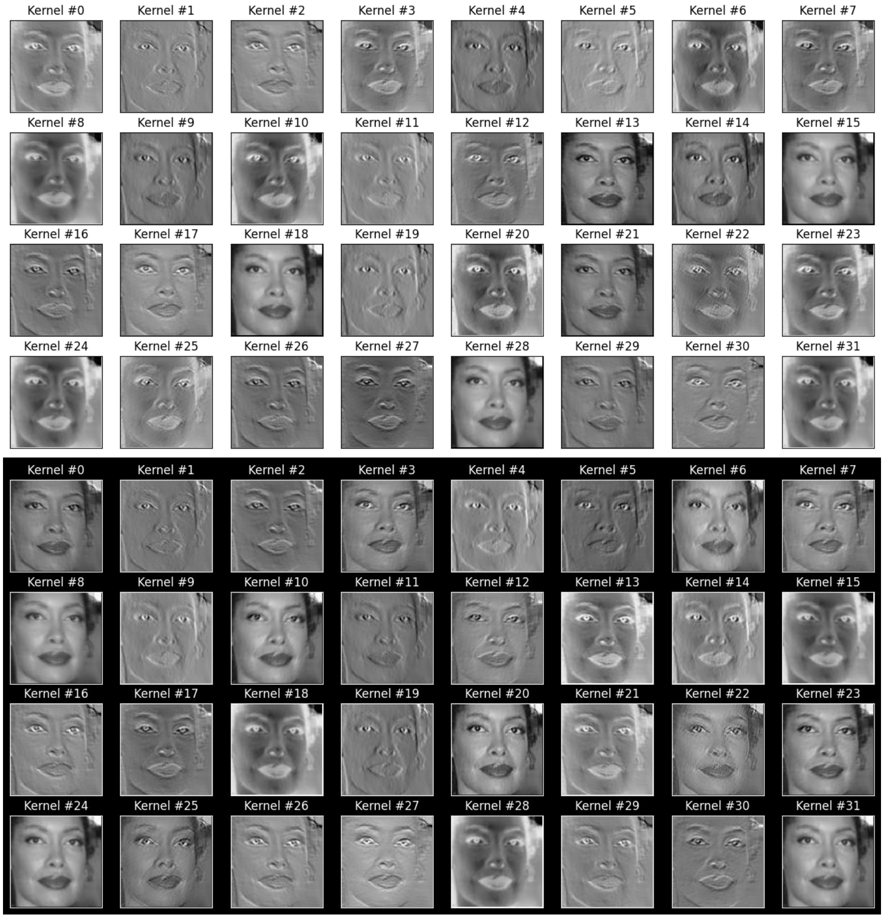


Figure 4.4. Comparison between leaked images and their negative image

4.2 Model Correlations

The correlations calculated between the 32 filters of the first layer of the models are presented below:

Modelo 1 - 128x128

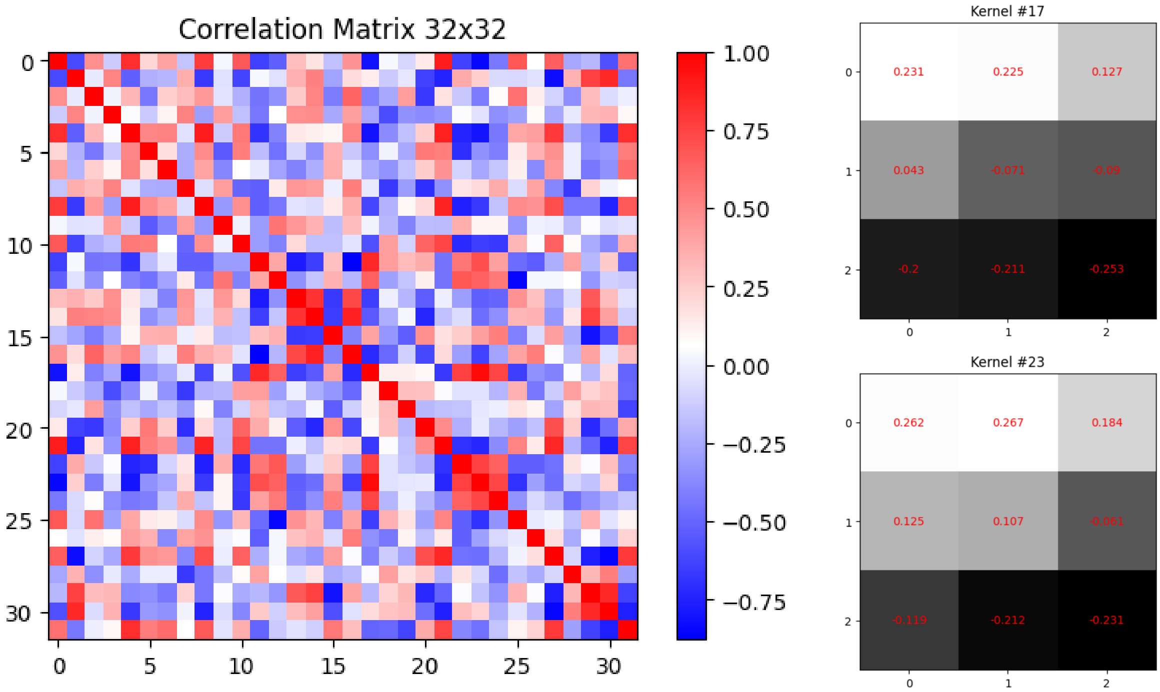


Figure 4.5. Highest correlation: 0.96 between kernel 17 and 23

Model 1.1 - 128x128 (database modification)

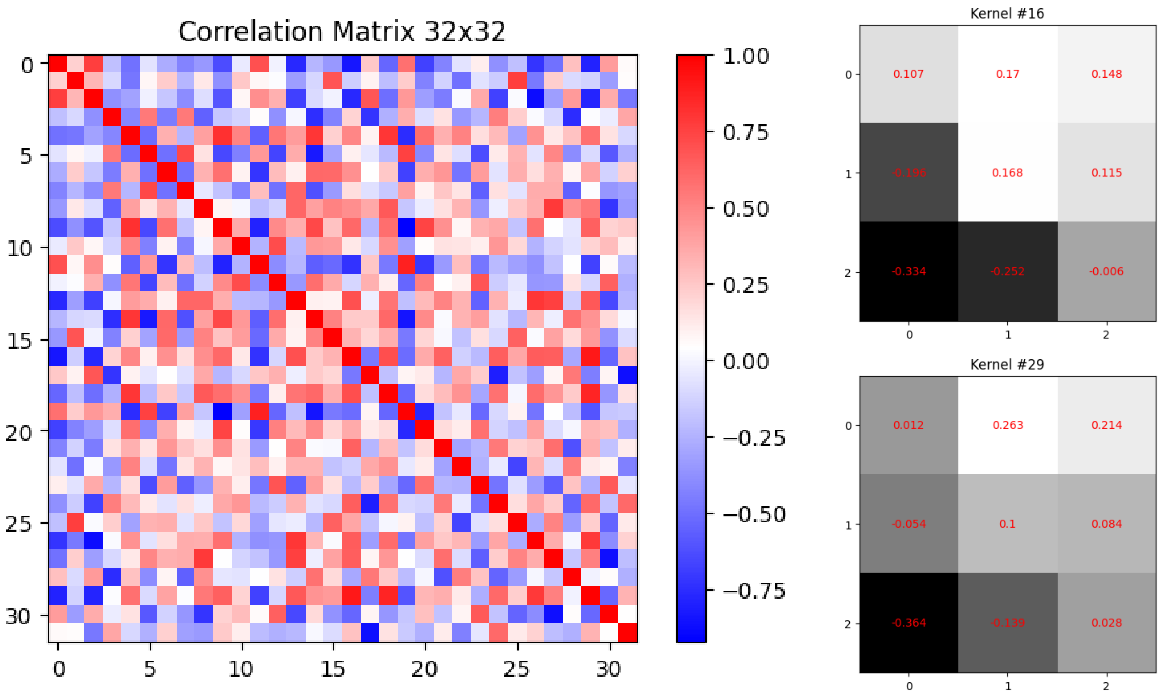


Figure 4.6. Highest correlation: 0.906 between kernel 16 and 29

Modelo 2 - 192x192

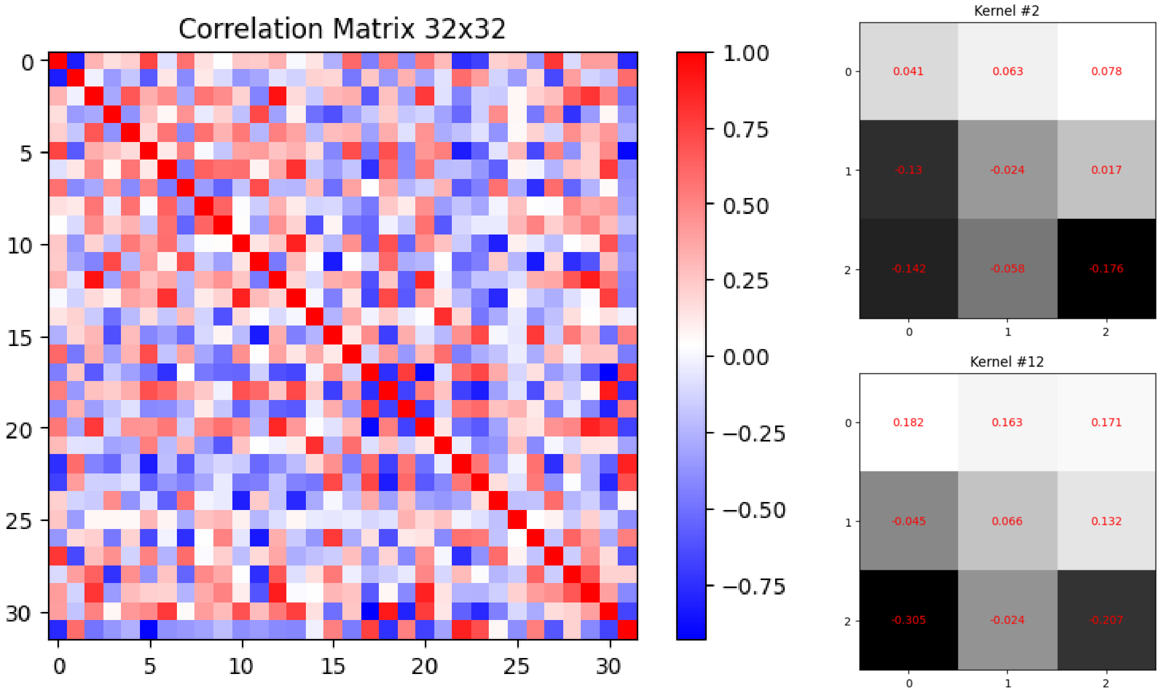


Figure 4.7. Highest correlation: 0.926 between kernel 2 and 12

By examining the correlation matrices and visualizing the first layer kernels in different models, we notice that while the filters may differ in their exact values, the relative relationship between the filters in terms of similarity or correlation remains consistent. This implies that despite variations in the network architecture or input dimensions, there is an underlying structure in the relationship between the filters that is maintained.

This consistency in the relationships between the filters suggests the presence of feature extraction patterns that are robust across different model configurations and input dimensions. These patterns can be of great importance in optimizing learning processes, reducing training times and improving the depth of learning in future CNN developments.

4.3 Infinity Norm

Modelo 1 - 128x128

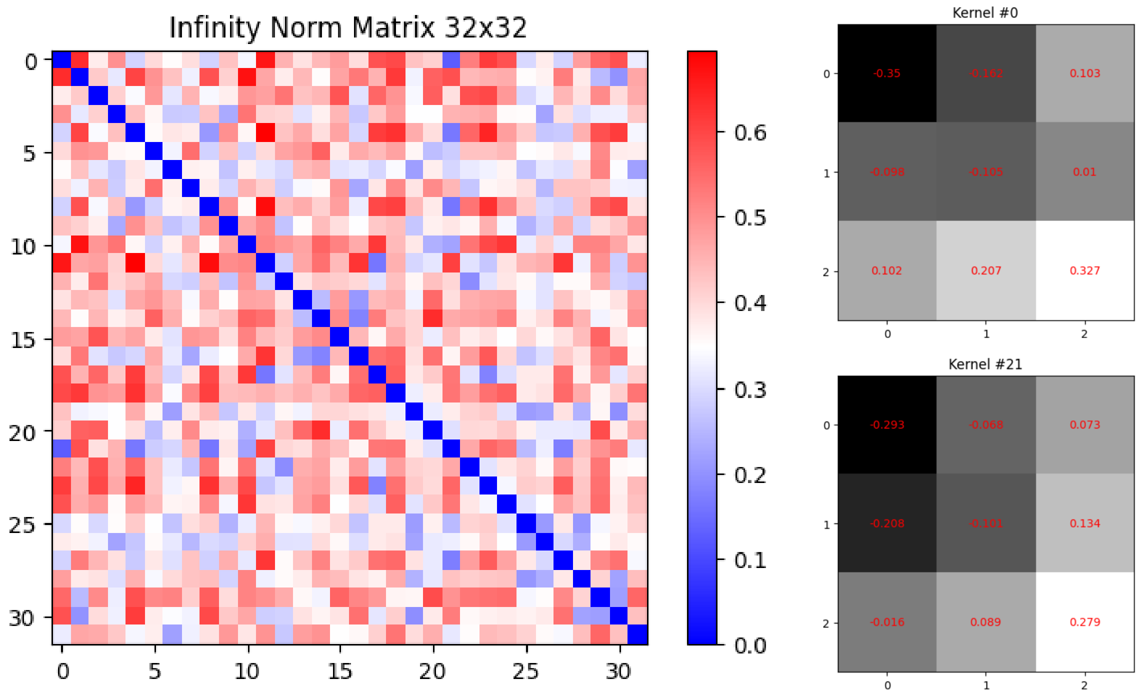


Figure 4.8. Shortest Distance: 0.124 between kernel 0 and 21

Model 1.1 - 128x128 (database modification)

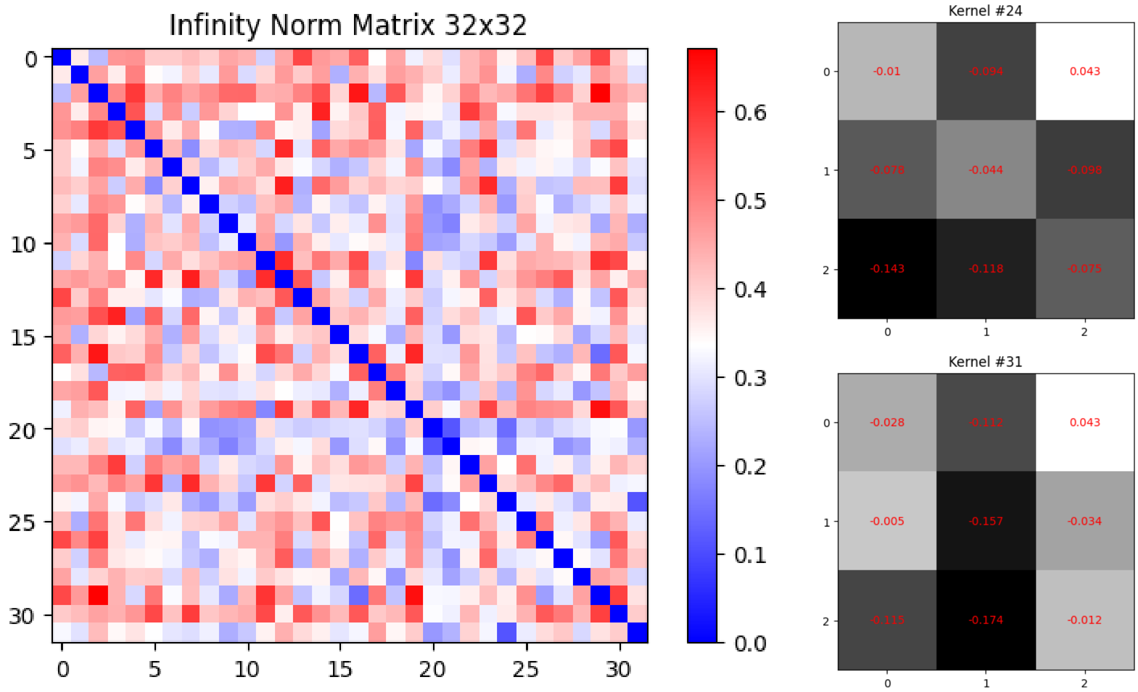


Figure 4.9. Shortest Distance: 0.112 between kernel 24 and 31

Modelo 2 - 192x192

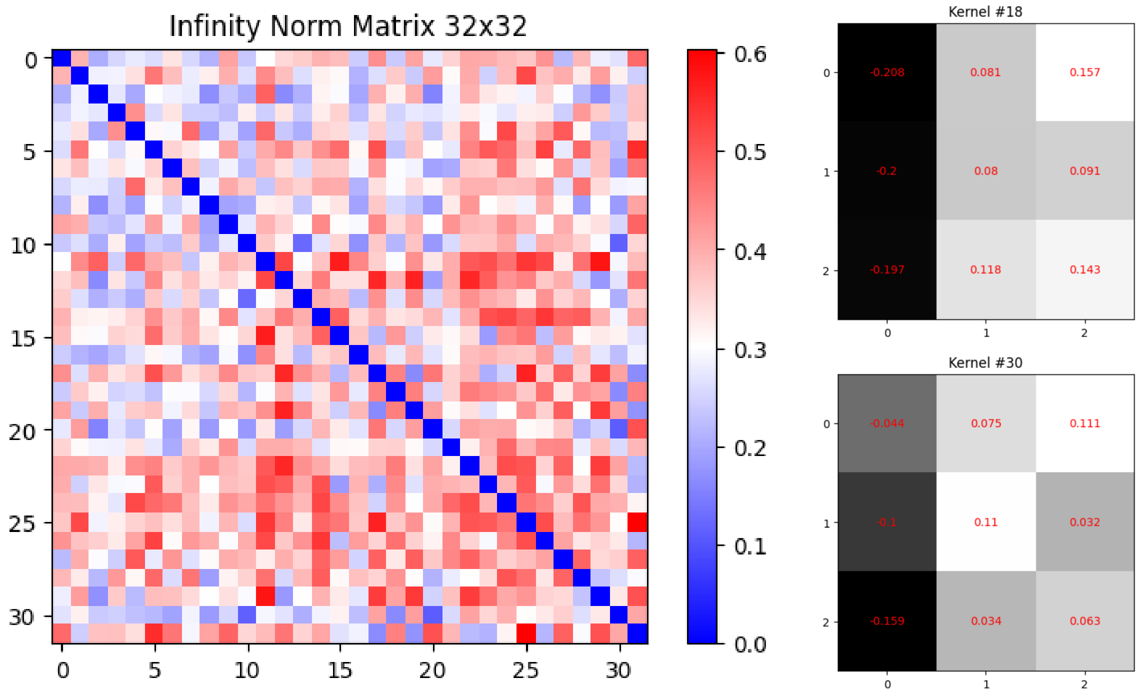


Figure 4.10. Shortest Distance: 0.164 between kernel 18 and 30

4.4 Cosine Similarity

Modelo 1 - 128x128

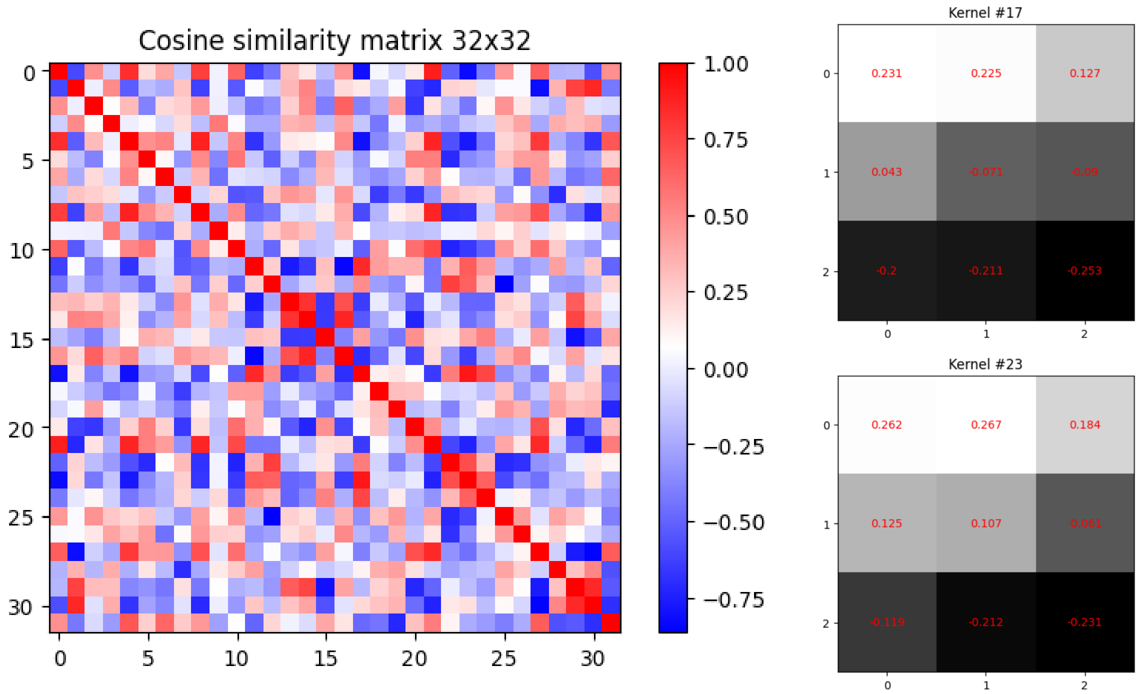


Figure 4.11. similarity: 0.914 between kernel 17 and 23

Model 1.1 - 128x128 (database modification)

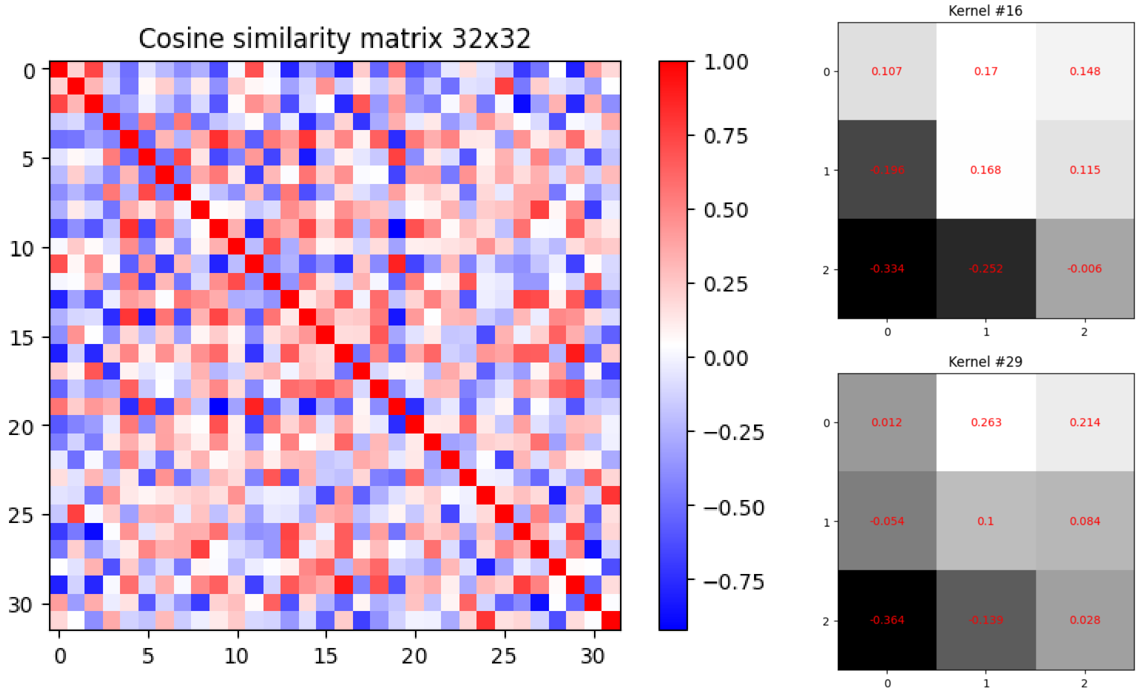


Figure 4.12. similarity: 0.897 between kernel 16 and 29

Modelo 2 - 192x192

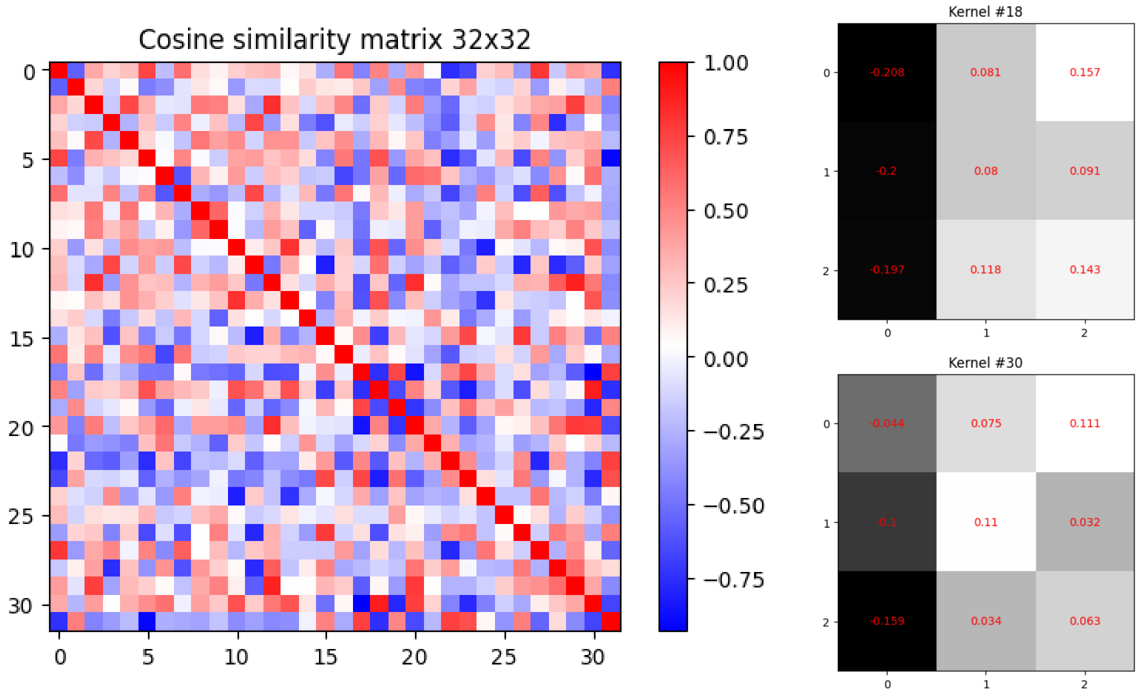


Figure 4.13. similarity: 0.890 between kernel 18 and 30

So far, we have explored the relationships between the most similar kernels in the CNN models. Now, we focus on identifying the opposites. This will allow us to better understand how CNN processes images and extracts distinctive features.

Modelo 1 - 128x128

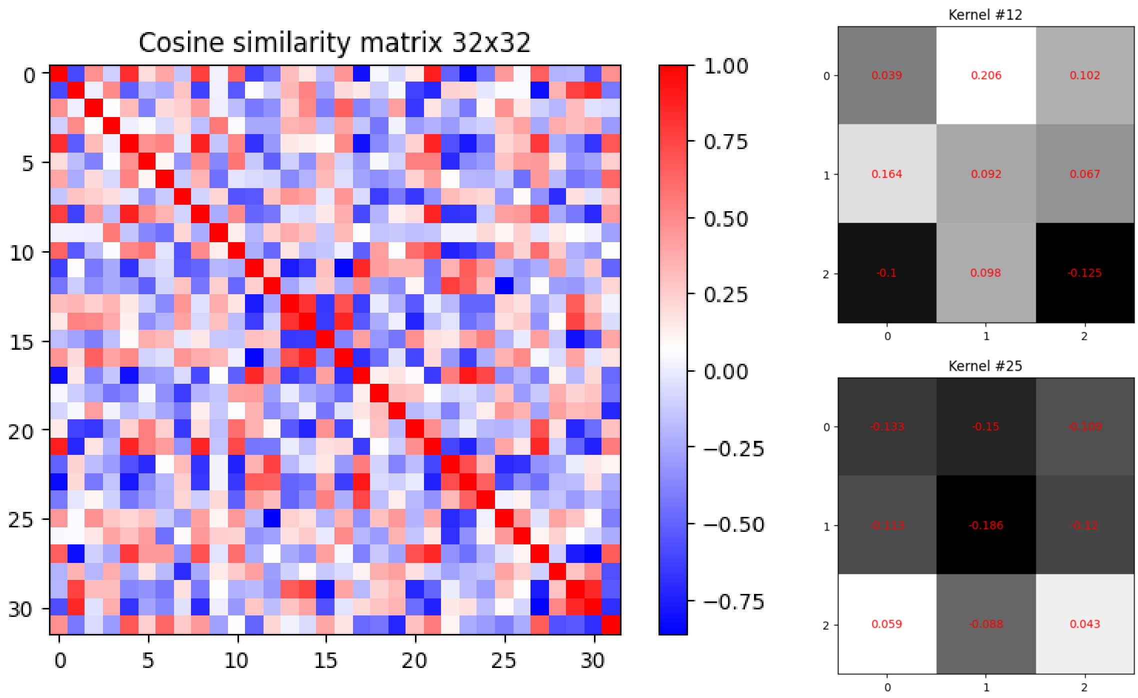


Figure 4.14. similarity: -0.864 between kernel 12 and 25

Model 1.1 - 128x128 (database modification)

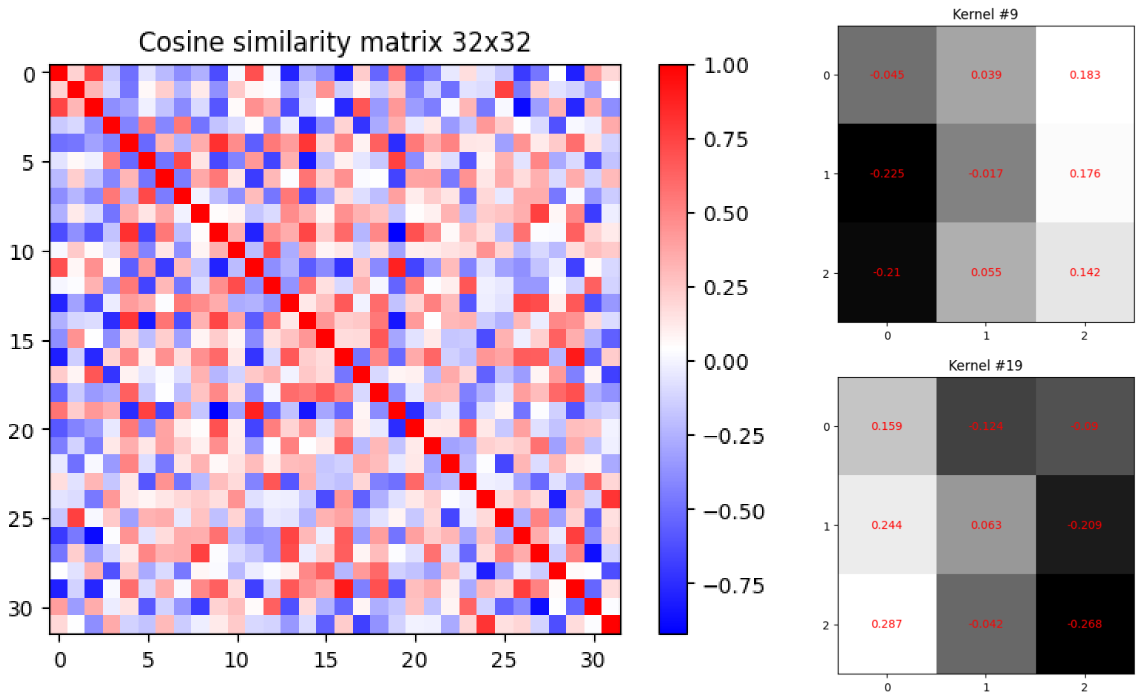


Figure 4.15. similarity: -0.920 between kernel 9 and 19

Modelo 2 - 192x192

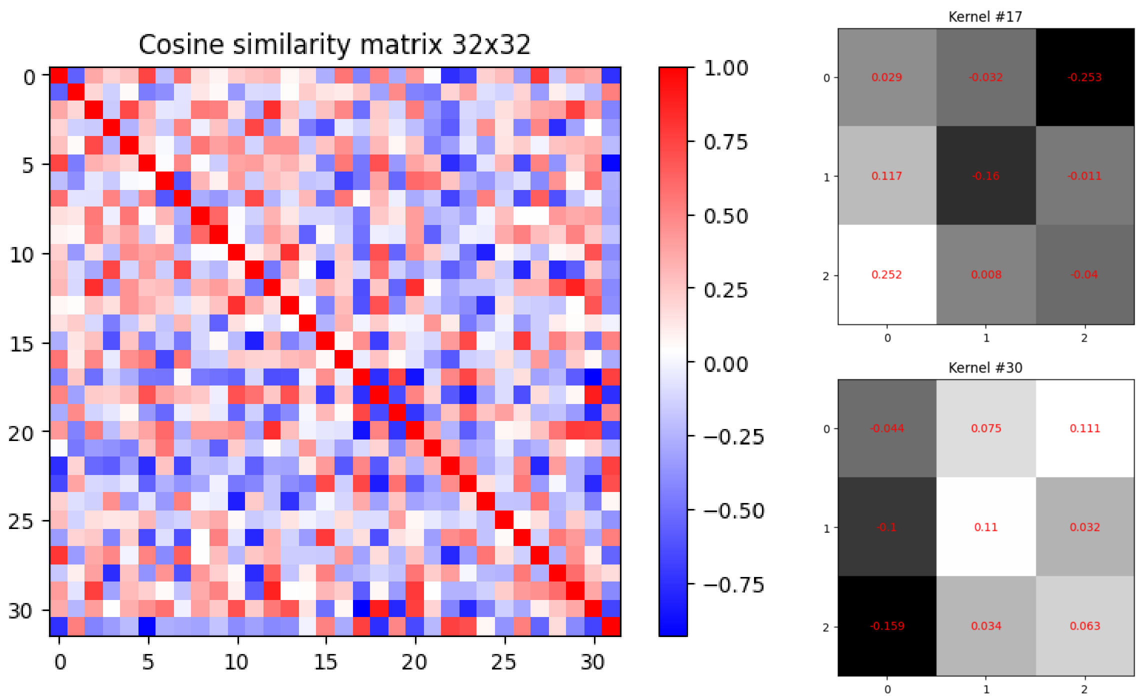


Figure 4.16. similarity: -0.927 between kernel 17 and 30

5 Conclusions

Throughout this document we investigated the performance and internal relationships of convolutional neural network (CNN) models for predicting age and gender from facial images using the UTKFace database. Our analysis revealed several key aspects:

Different input sizes: We trained three CNN models with different input sizes (128x128, 192x192) and database modifications. This allowed comparing and evaluating the performance of models in different configurations.

we study the relations between filters in the original layers of the model. We observe that even in structural changes in training input, the relative relationships remain the same, suggesting the underlying patterns of feature extraction.

We calculate the Infinity Norm matrix to evaluate the differences between the kernels of the first layer. This measure provided information about the magnitude of the differences and the convergence of features in the models, also we observed remarkable similarities between the filtered images and found that by reversing the channels of some images, we were getting exact matches with others. This highlights the complexity of the relationships between the filters and the information in the images.

The same analyzes were made for the last layer and as expected many of the calculations carried out by the network lead to the same point, this is why the filters are the same, we can see this as soon as we increase the possible filters, in the analysis of 64x64 A.1, 128x128 A.2 and 256x256 A.3 in appendix .

To end, our results show that, despite differences in model configurations and input dimensions, there is consistency for the original CNN layer filters. These observations are very important to optimize the learning process, reduce training time and improve the depth of learning in the future development of CNNs.

6 Bibliography

- Wikipedia contributors. (2023a). Uniform norm. Wikipedia. https://en.wikipedia.org/wiki/Uniform_norm
- Wikipedia contributors. (2023b). Correlation. Wikipedia. https://en.wikipedia.org/wiki/CorrelationCorrelation_matrices
- Wolfram Research, Inc. (s. f.). Wolfram Search. https://mathworld.wolfram.com/search/?q=Infinity+Norm

APPENDIX

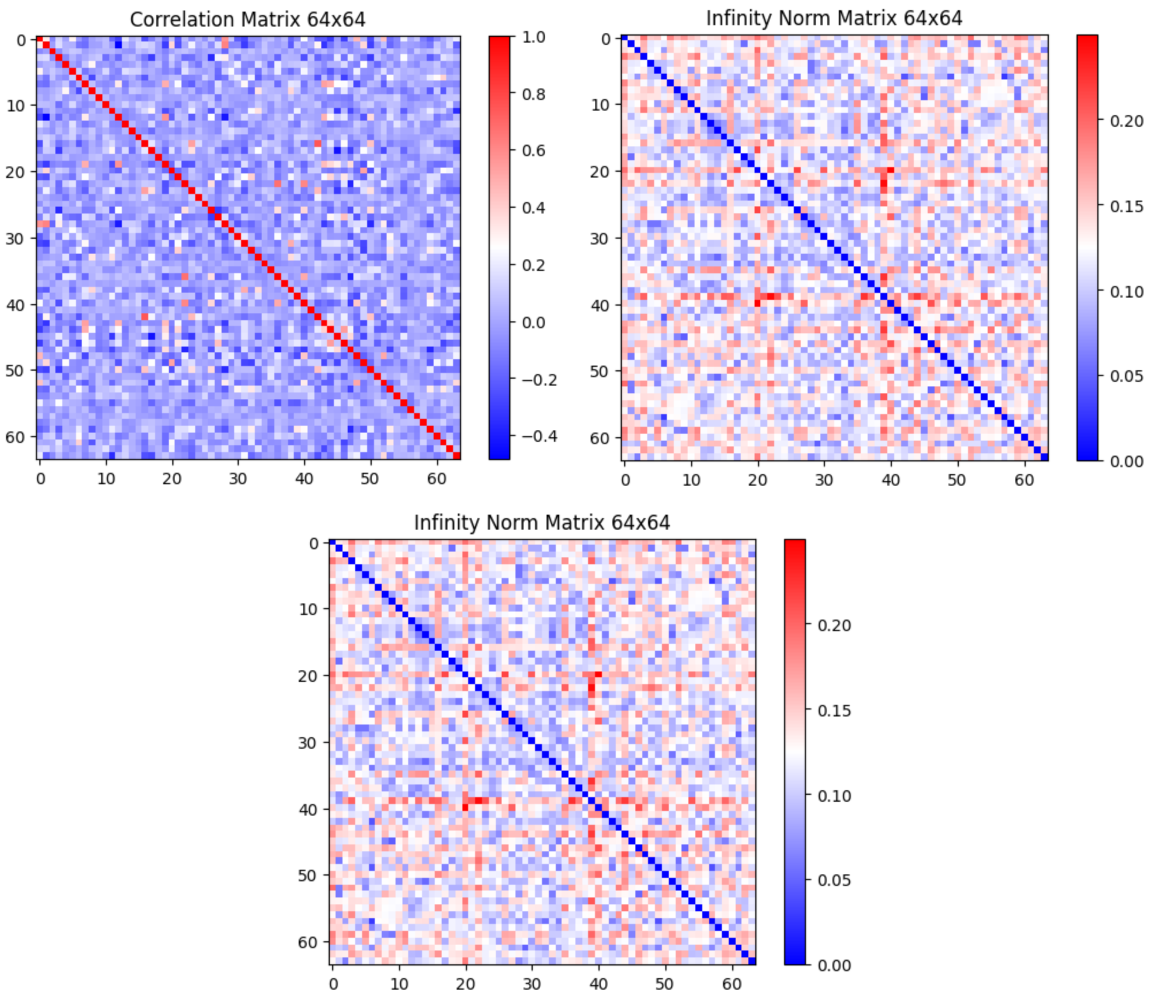


Figure A.1. Analysis for the 64x64 filter layer

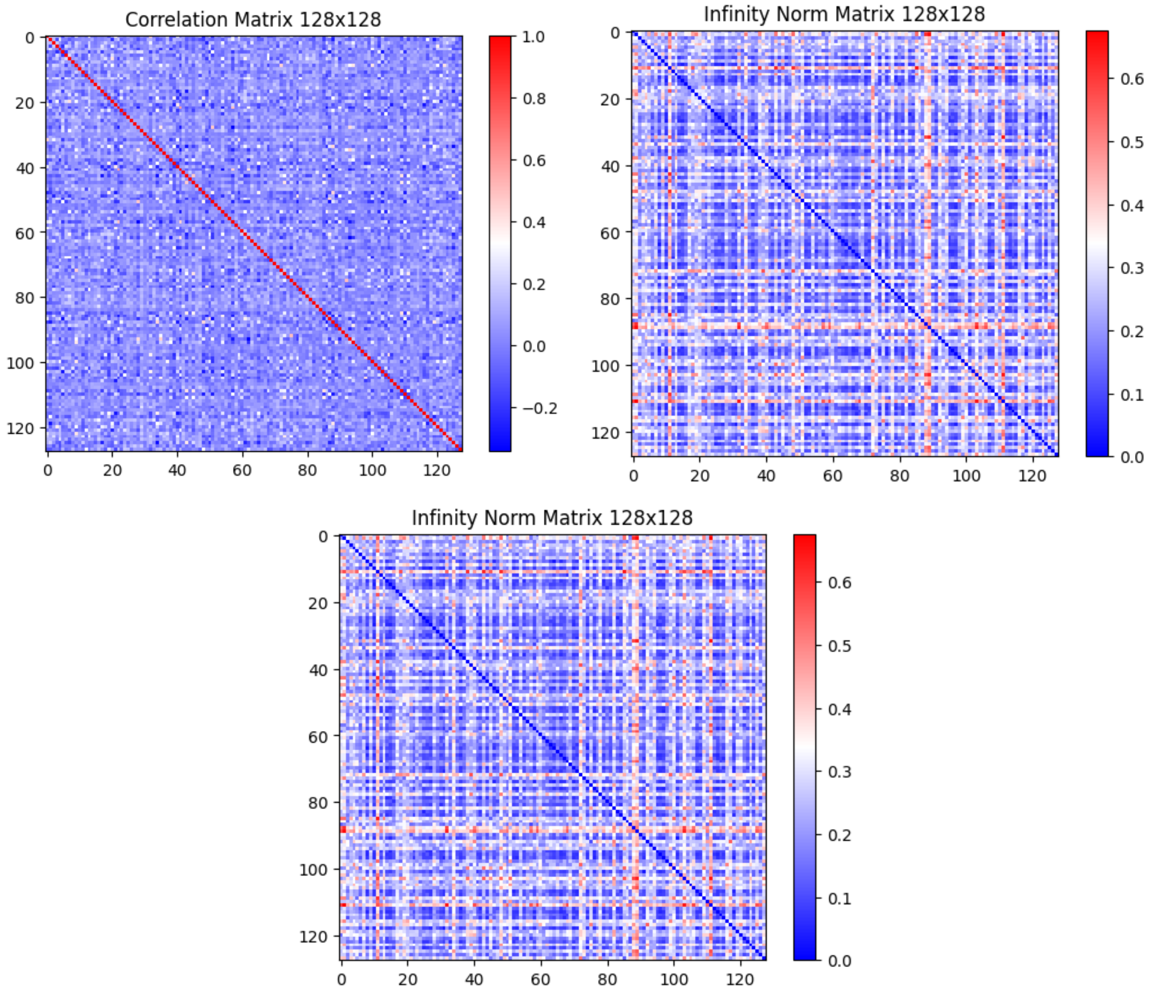


Figure A.2. Analysis for the 128x128 filter layer

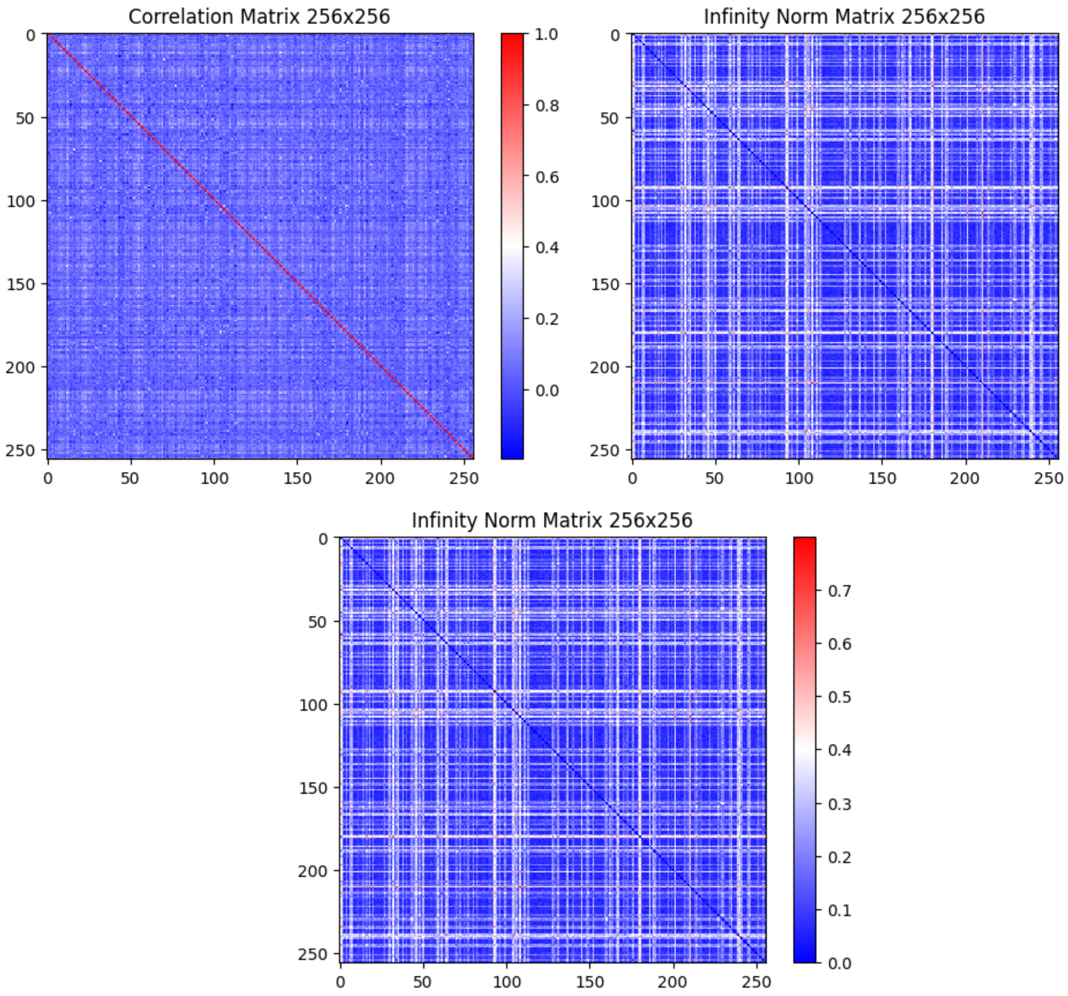
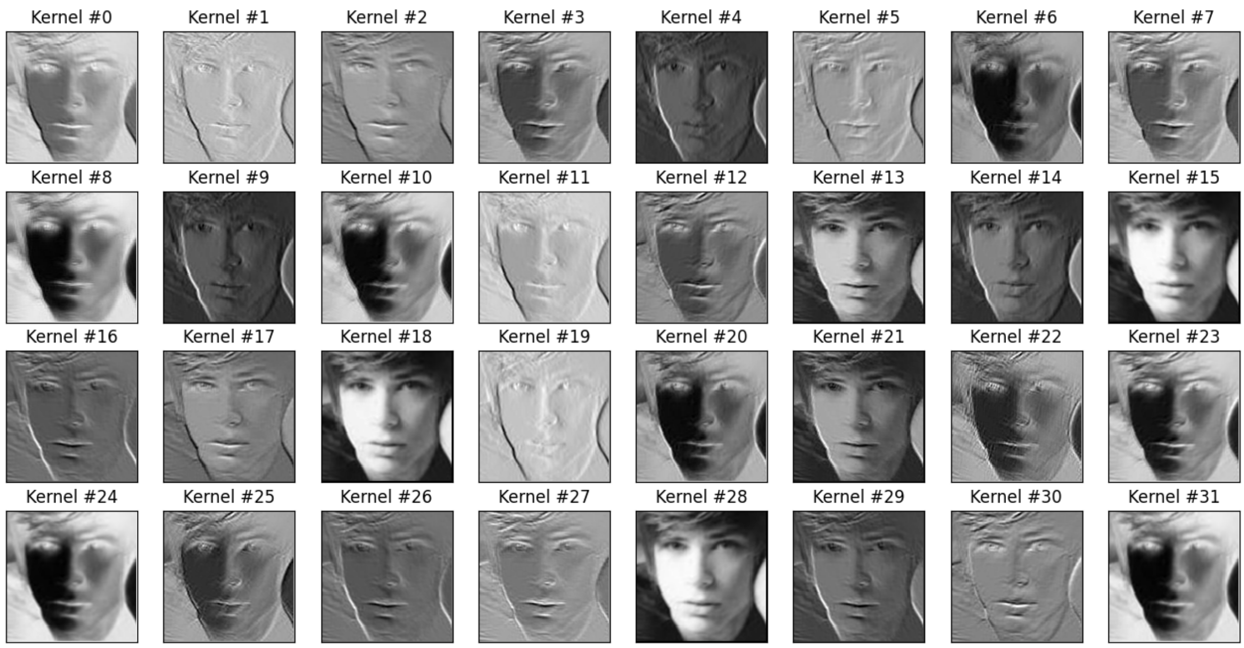
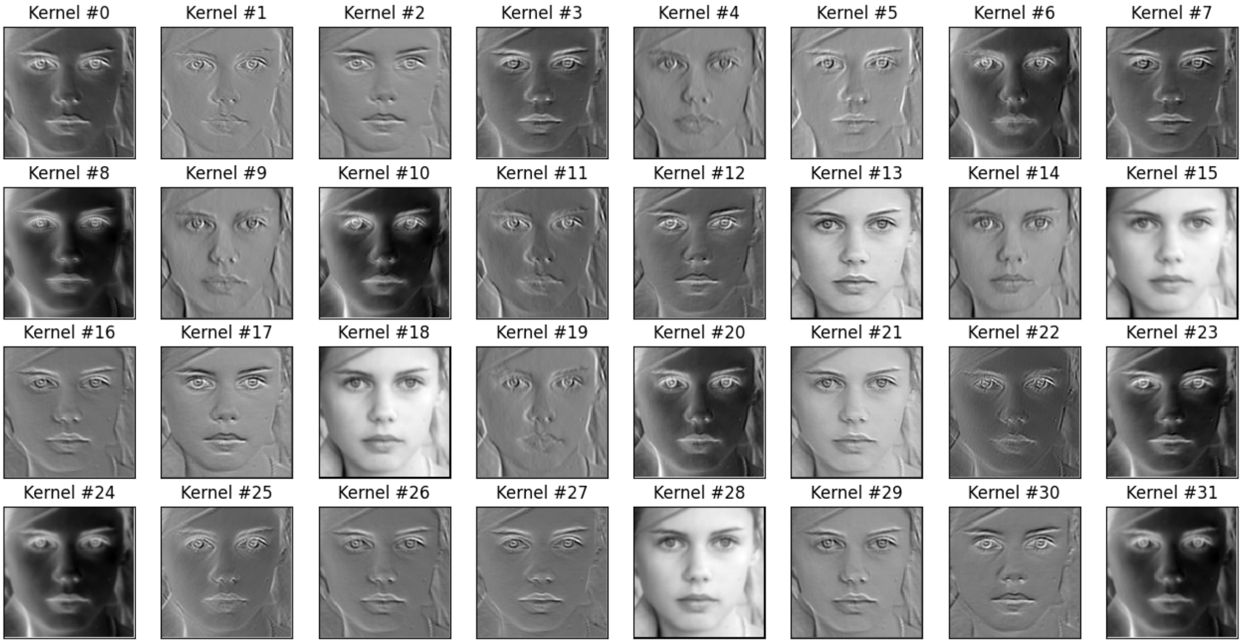


Figure A.3. Analysis for the 256x256 filter layer




Kernel #0
Kernel #1
Kernel #2
Kernel #3
Kernel #4
Kernel #5
Kernel #6
Kernel #7
Kernel #8
Kernel #9
Kernel #10
Kernel #11
Kernel #12
Kernel #13
Kernel #14
Kernel #15
Kernel #16
Kernel #17
Kernel #18
Kernel #19
Kernel #20
Kernel #21
Kernel #22
Kernel #23
Kernel #24
Kernel #25
Kernel #26
Kernel #27
Kernel #28
Kernel #29
Kernel #30
Kernel #31