HOTDOG / NOT HOTDOG: A DEEP LEARNING APPROACH

Samer Bujana, Rasmus Clausen, Alba Gonzalo, Robert Jarvi, Leonardo Rodovero

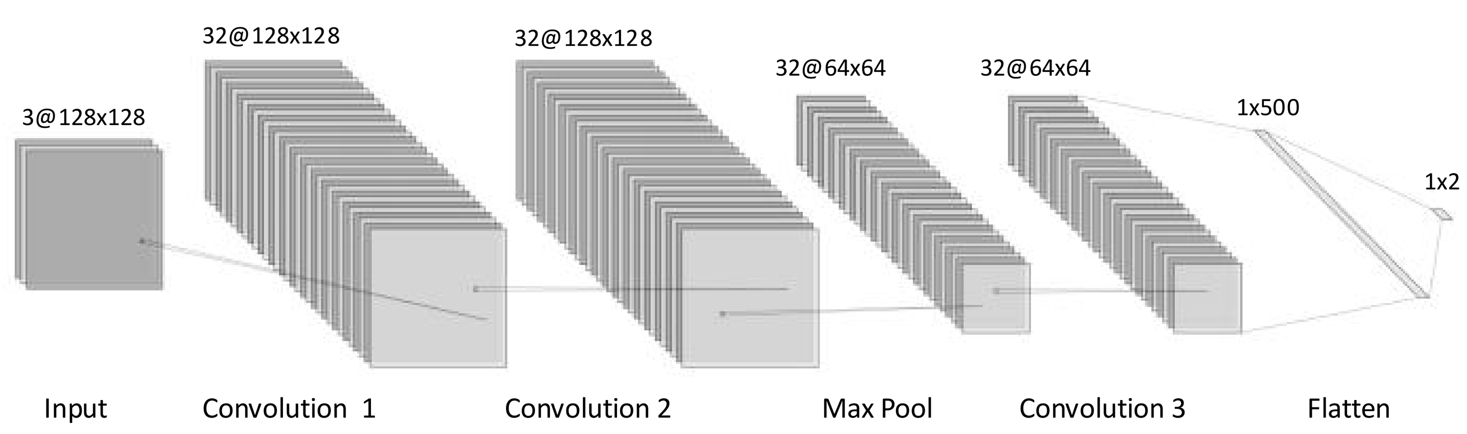
DTU

INTRODUCTION

The purpose of this project is to develop a CNN that is capable of identifying the presence of hotdogs in images. This model has been designed from scratch, starting first with a very basic architecture that has been increased in complexity through different iterations. First, we tried different methods to improve the performance of the network, and then, we combined them to see if these boosts in performance could be complementary.

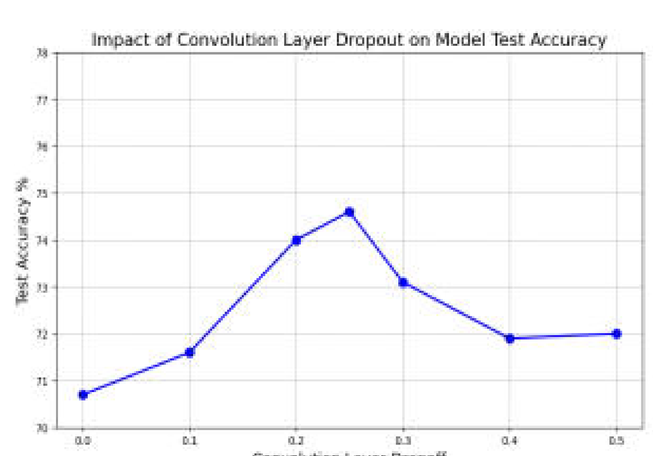
PRELIMINARY TESTING

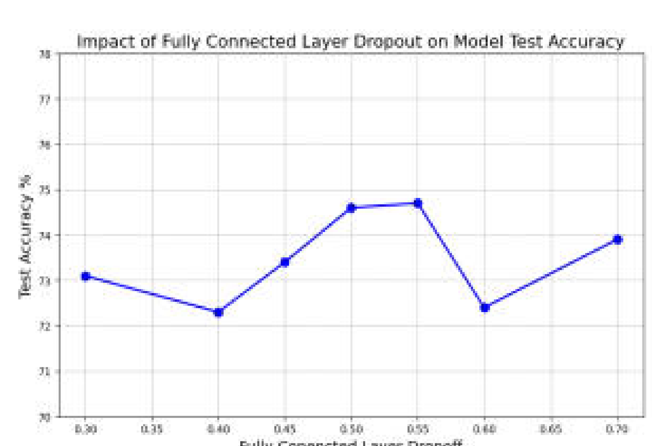
To compare the benefits and performance of various network architecture components, we created a simple network as a baseline.



DROPOUT OPTIMIZATION

We tested dropout layers in both convolutional and fully connected layers separately. We found that **p=0.25** for convolutional layers and **p=0.55** achieved optimal accuracy, in line with existing research [1].

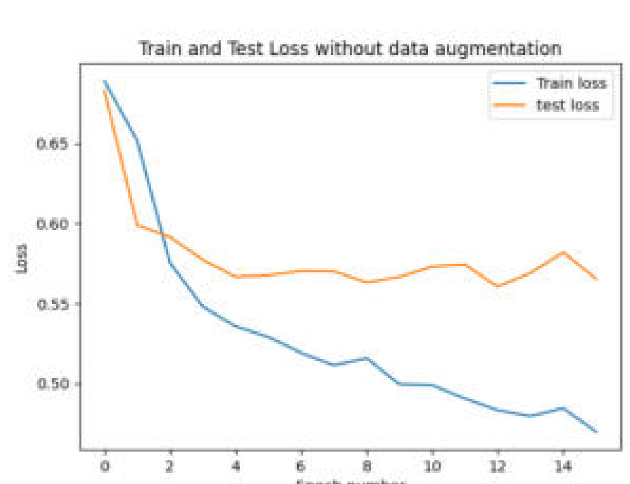


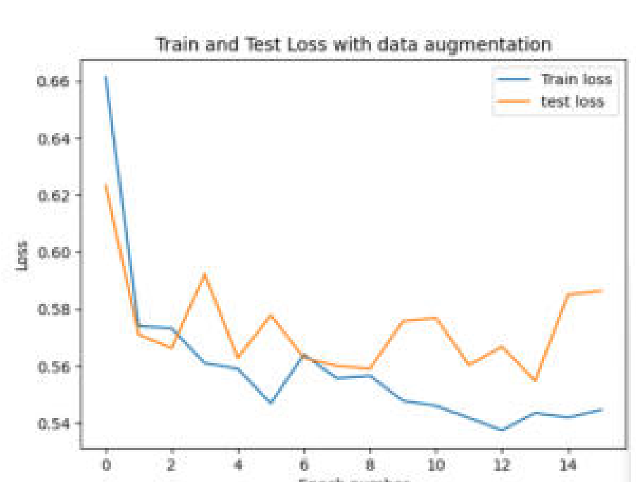


DATA AUGMENTATION

Data augmentation didn't improve accuracy, but it helped reduce existing overfitting. We implemented different combinations of:

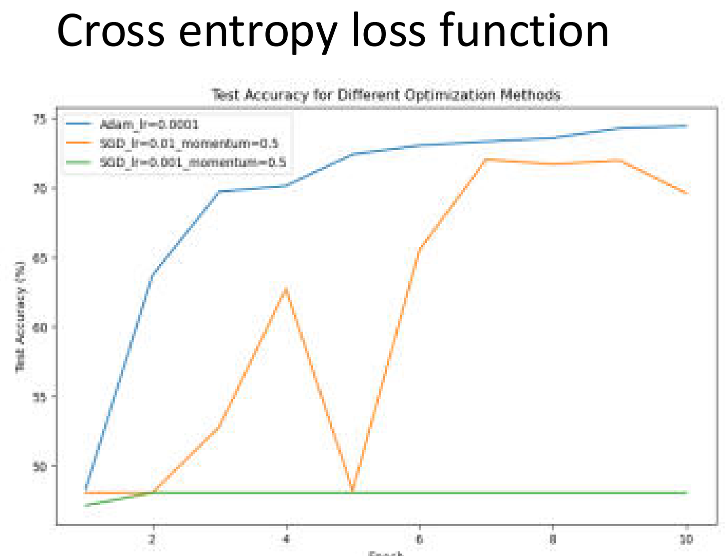
- Rotation
- Horizontal/ Vertical Flip
- Normalization
- Brightness, contrast, saturation (ColorJitter)
- Affine
- Zoom



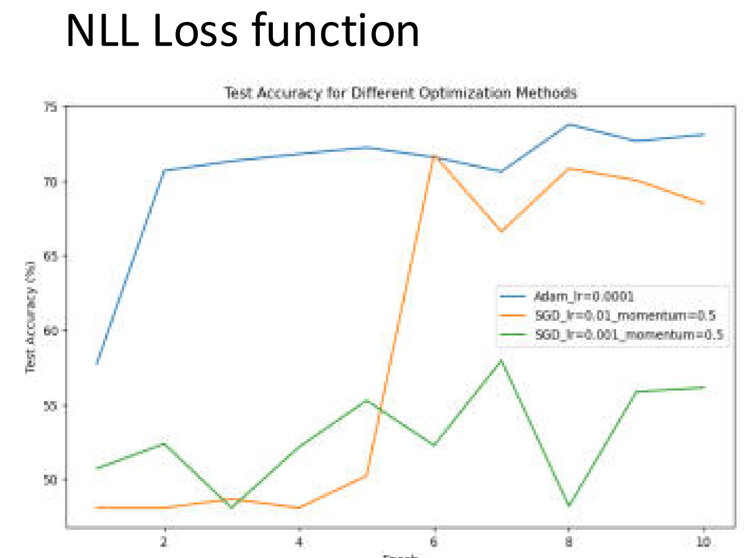


LOSS FUNCTION OPTIMIZATION

Cross entropy loss function

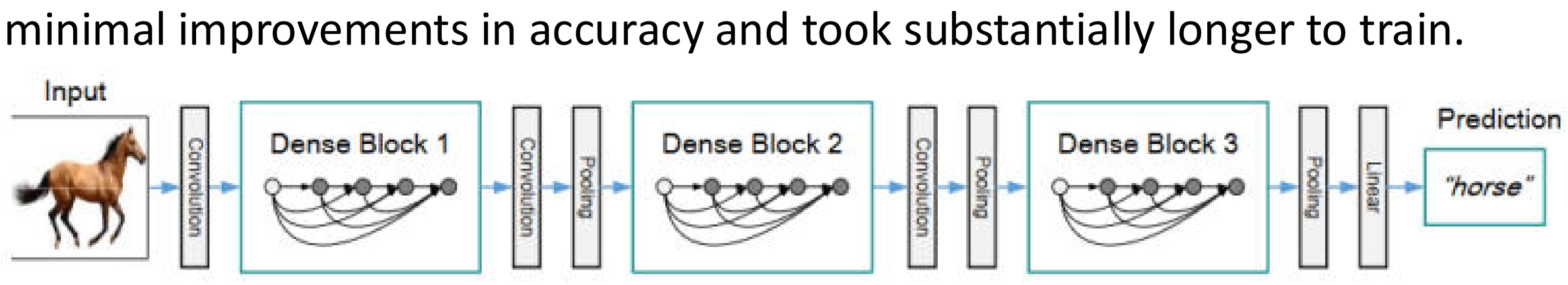


NLL Loss function



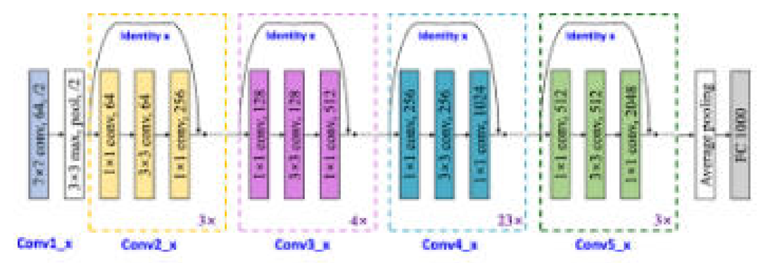
DENSENET

To try and increase the test accuracy we recreated a DenseNet model with DenseNet 121, with configuration [6,12,24,16] [2]. This configuration had minimal improvements in accuracy and took substantially longer to train.



TRANSFER LEARNING COMPARISON

We used a **pretrained ResNet model** on the Hot Dog dataset to serve as a baseline for optimal classification performance. The pre-trained model achieved an accuracy of **90.4%.**



TESTED MODELS AND PERFORMANCE

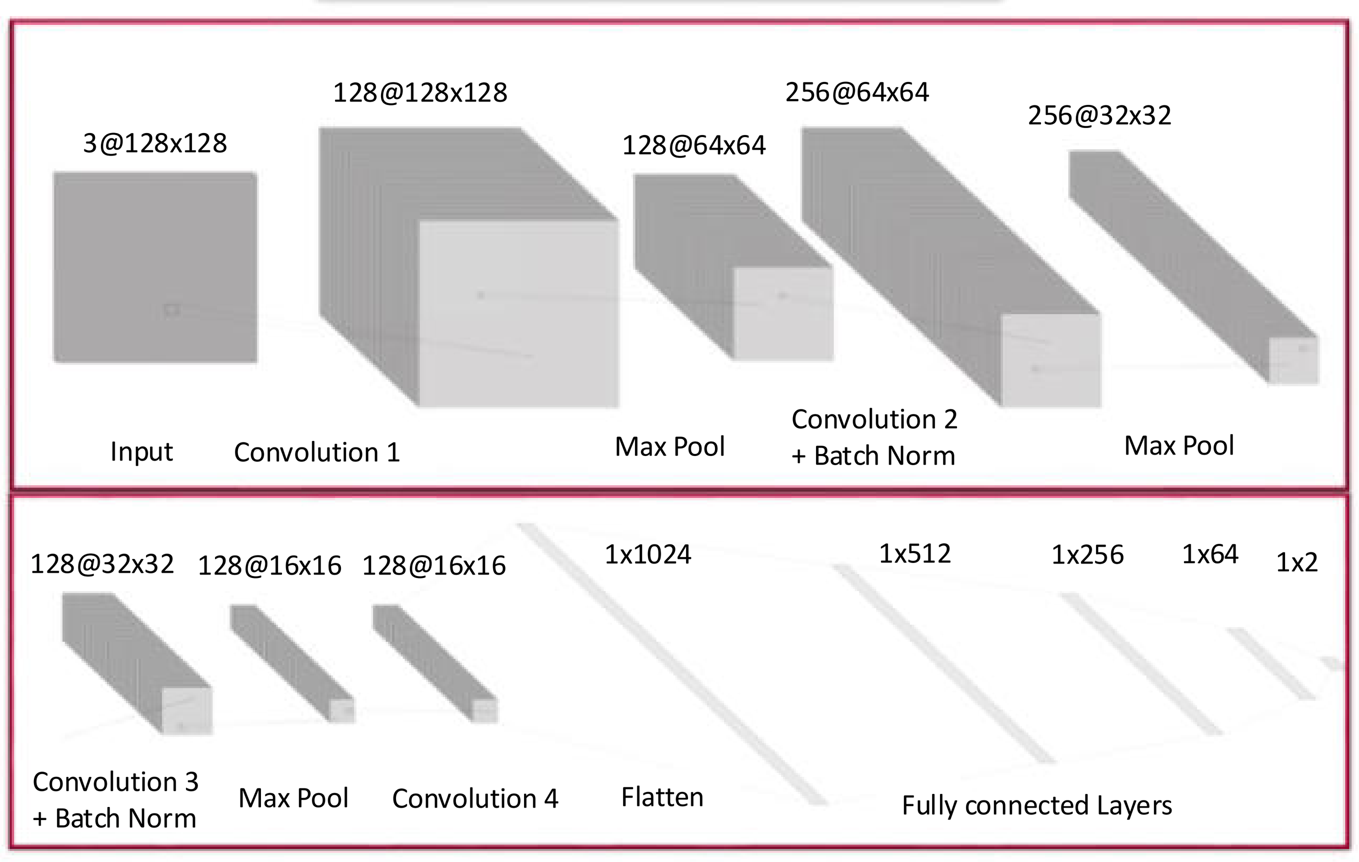
Each model was trained with the following parameters:

- Epochs = 16
- Cross entropy function
- SGD
- Momentum = 0.9 & lr=0.1

Model	Normalization	Dropout	augumentation	%accuracy
Baseline test 0	✗	✗	✗	72.2%
Baseline test 2	●	✗	✗	76%
Baseline test 3	✗	●	✗	76,4%
Baseline test 4	●	●	✗	77%
Baseline test 5	●	●	●	77%
Best model	●	●	●	81.1%

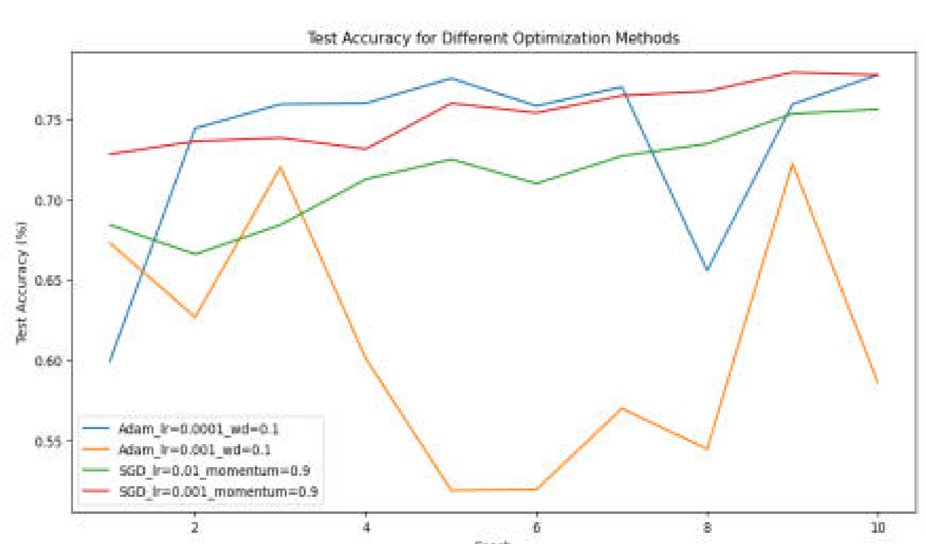
FINAL MODEL

Test accuracy over 80% !

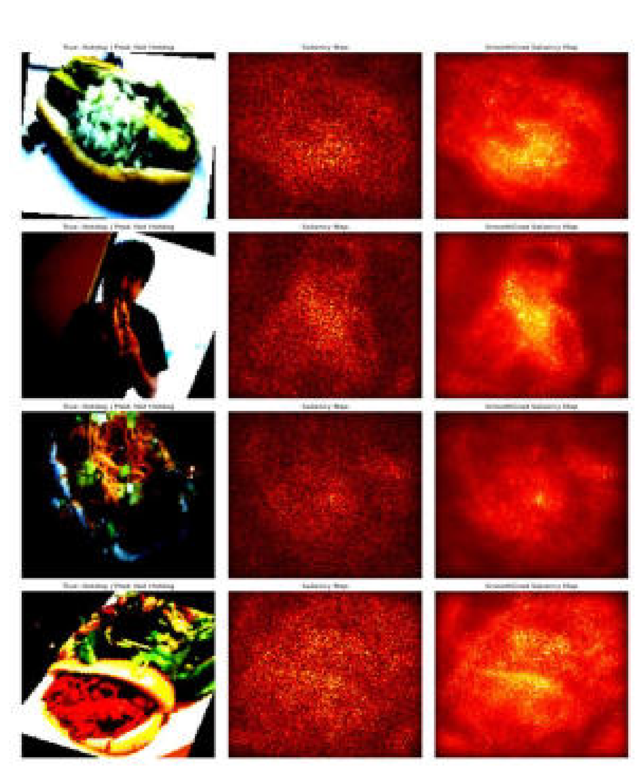


FINAL RESULTS ON THE MODEL'S PERFORMANCE

After extensive tests and adjustments, we achieved a model with **81.1% test accuracy** using the **cross-entropy** loss function. Our test loss graph demonstrates our architecture effectively avoided overfitting.



Our analysis of misclassified images and **saliency maps*** revealed that the model struggles to identify hot dogs when **people** are present. We believe that the **wide variety of hot dog types** in the dataset is not sufficiently represented, contributing to this challenge.



REFERENCES AND USE OF AI *

[1] http://mipal.snu.ac.kr/images/1/16/Dropout_ACCV2016.pdf, [2] https://paperswithcode.com/paper/densely-connected-convolutional-networks

AI tools were used to aid in code generation and debugging, and to help with understanding of DL topics.