

**NICOLI PINHEIRO DE ARAUJO**

**ESTIMAÇÃO INTELIGENTE DE IDADE DE TELESPECTADORES PARA  
APLICAÇÕES DE SUGESTÃO DE CONTEÚDO EM *SMART* TVS**

Trabalho de Conclusão de Curso apresentado  
à banca avaliadora do Curso de Engenharia  
de Computação, da Escola Superior de  
Tecnologia, da Universidade do Estado do  
Amazonas, como pré-requisito para obtenção  
do título de Engenheira de Computação.

Orientador(a): Profa. Dra. Elloá Barreto Guedes da Costa

Manaus – Novembro – 2018

# Capítulo 1

## Resultados e Discussão

Considerando a estratégia descrita na solução proposta, os resultados da execução das CNNs aplicadas ao problema de estimação de idade a partir de uma imagem de face são apresentados a seguir. Estes resultados estão organizados segundo abordagens sequenciais, que contemplam desde as técnicas mais elementares, e que vão aumentando o grau de complexidade conforme uso de estratégias específicas da prática de DL para a resolução de problemas práticos.

### 1.1 Abordagem 1: LeNet e AlexNet com Imagens Normalizadas

A primeira abordagem de treinamento considerou o uso dos modelos de maneira canônica, isto é, tais como são definidos na literatura. Adotou-se as funções de ativação não-lineares *ReLU* e *Leaky ReLU* por serem simples de calcular e por satisfazerem os critérios de continuidade e diferenciação, requeridos pelo algoritmo de *backpropagation*, conforme discutido anteriormente na Seção ??.

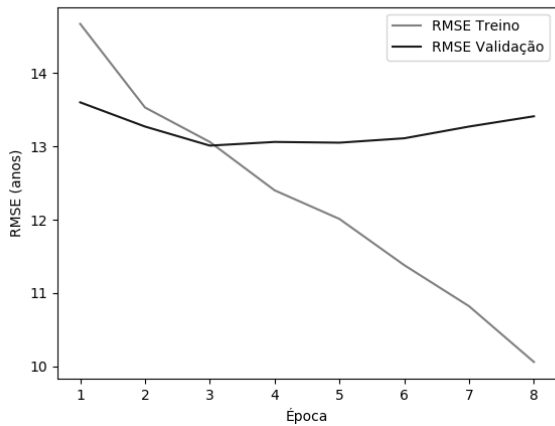
As imagens da base de dados foram normalizadas antes de serem apresentadas às redes. Todos os valores dos pixels componentes das imagens foram escalonados para o intervalo  $[0, 1]$  por meio de uma divisão por 255. A prévia normalização das imagens antes da apresentação às CNNs colabora para uma melhor execução do gradiente descendente e diminui a variância

nos pesos.

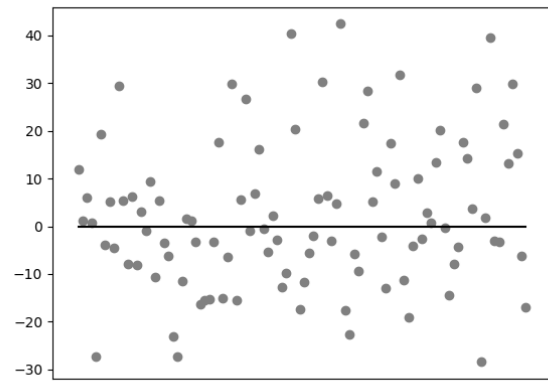
Os treinamentos destas duas arquiteturas duraram aproximadamente 16 e 12 horas respectivamente, em uma instância do Google Compute Engine com 4 CPUs virtuais e 15 GB de RAM. Os gráficos de treinamento e as retas zero obtidas a partir da apresentação do conjunto de teste aos modelos consolidados podem ser vistos na Figura 1.1. É possível notar que ambas as redes sofreram *overfitting* e obtiveram grande margem de erro.

Figura 1.1: Resultados do treinamento e teste da CNN LeNet de acordo com a Abordagem 1.

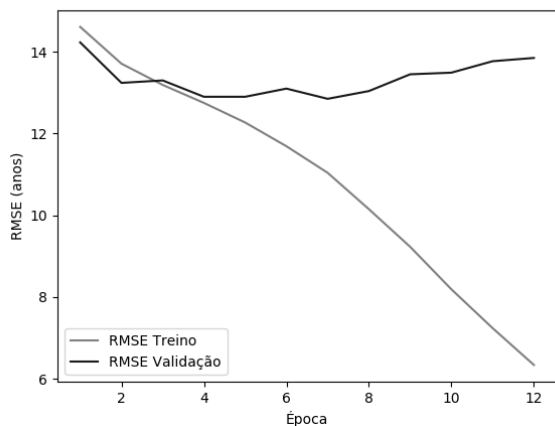
(a) RMSE de treinamento da arquitetura LeNet utilizando funções de ativação *ReLU*.



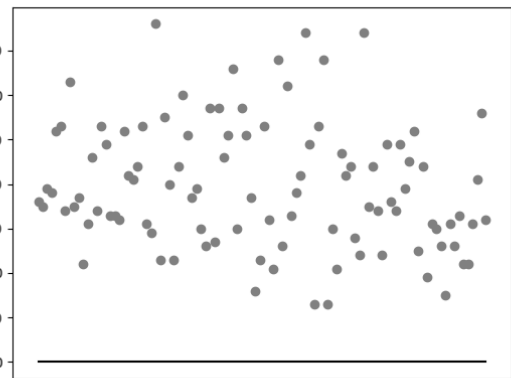
(b) Reta-0 LeNet *ReLU*.



(c) RMSE de treinamento da arquitetura LeNet utilizando funções de ativação *Leaky ReLU*.



(d) Reta-0 LeNet *Leaky ReLU*.



Obedecendo ao método de validação cruzada *holdout* previamente mencionado, os resultados desta abordagem encontram-se sintetizados na Tabela 1.1. É possível constatar que as CNNs

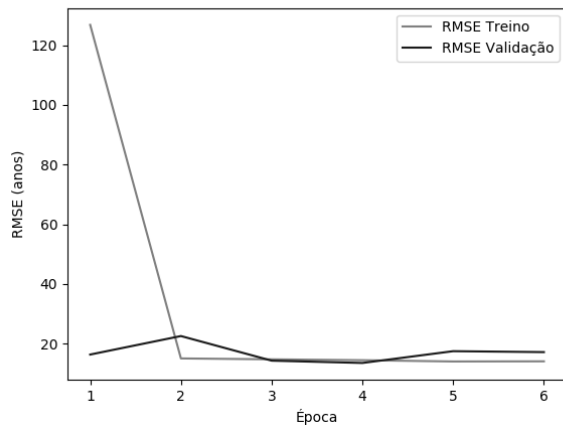
com função de ativação *ReLU* obtiveram melhor desempenho, com a arquitetura LeNet, em particular, com resultados ligeiramente superiores.

Tabela 1.1: Resultados do treino e teste dos modelos propostos na Abordagem 1.

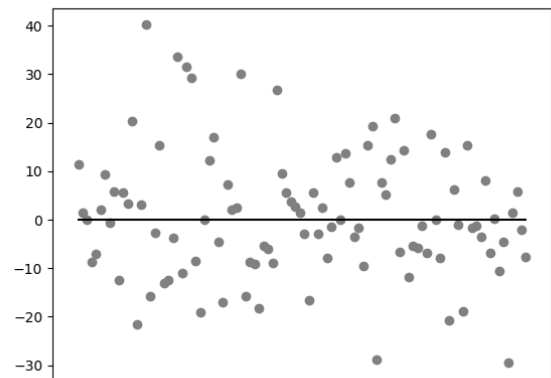
Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
LeNet	<i>ReLU</i>	4	10.53	13.55
LeNet	<i>Leaky ReLU</i>	8	38.33	40.82
AlexNet	<i>ReLU</i>	5	11.03	13.76
AlexNet	<i>Leaky ReLU</i>	5	39.27	41.97

Figura 1.2: Resultados do treinamento e teste da CNN AlexNet de acordo com a Abordagem 1.

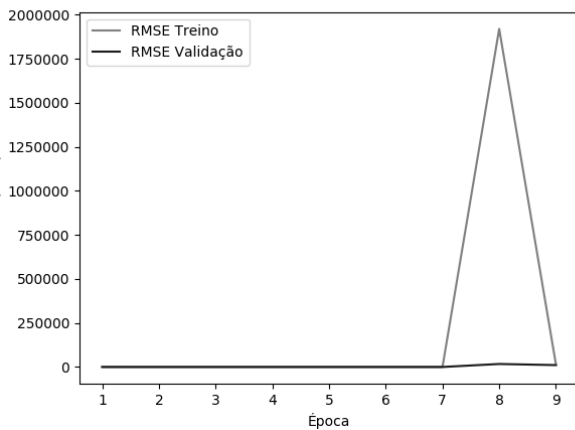
(a) Treinamento AlexNet *ReLU*.



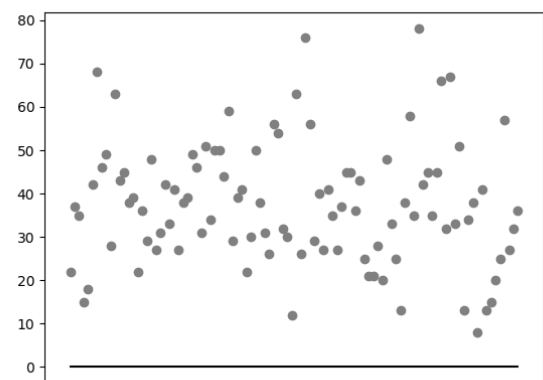
(b) Reta-0 AlexNet *ReLU*.



(c) Treinamento AlexNet *Leaky ReLU*.



(d) Reta-0 AlexNet *Leaky ReLU*.



## 1.2 Abordagem 2: Introduzindo *Data Augmentation*

A abordagem anterior consistiu essencialmente da utilização dos modelos tal como foram definidos e com uma simples operação de adequação dos dados de entrada por meio de normalização. Porém, em problemas de Visão Computacional, é comum aplicar técnicas de *data augmentation* com vistas a aumentar artificialmente o conjunto de dados, fazendo com que o modelo, em sua fase de treinamento, não seja exposto à mesma entrada em mais de uma ocasião. Isto previne *overfitting* e colabora para uma melhor generalização (CHOLLET, 2017).

As técnicas de *data augmentation* consideradas foram a rotação entre 0 e 20 graus no sentido horário ou anti-horário, zoom de 0.8 a 1.2 vezes, inversão horizontal com probabilidade de ocorrência de 0.5 ou translação com probabilidade igual a 0.2.

Os gráficos das métricas de desempenho coletadas durante o treinamento e a reta-0 obtida a partir dos dados de teste em cada uma destas quatro configurações são ilustrados nas Figuras 1.3 e 1.4.

De maneira análoga, as métricas de desempenho coletadas encontram-se detalhadas na Tabela 1.2. Nota-se que o número de épocas no treinamento foi maior que a abordagem anterior, indicando que houve um cenário mais favorável para o aprendizado dos padrões nos dados. De maneira geral, as métricas obtidas não fornecem uma evidência forte de que esta segunda abordagem produz resultados mais significativos que a primeira mas, no caso da CNN AlexNet com *ReLU*, os resultados foram comparáveis.

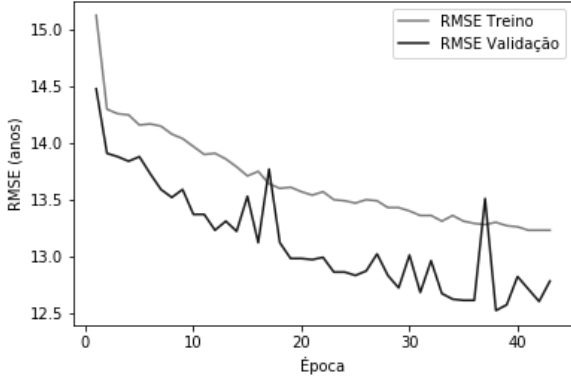
Tabela 1.2: Resultados do treino e teste dos modelos propostos na Abordagem 2.

Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
LeNet	<i>ReLU</i>	39	37.85	40.27
LeNet	<i>Leaky ReLU</i>	21	38.50	41.06
AlexNet	<i>ReLU</i>	16	11.59	14.59
AlexNet	<i>Leaky ReLU</i>	16	28.06	31.81

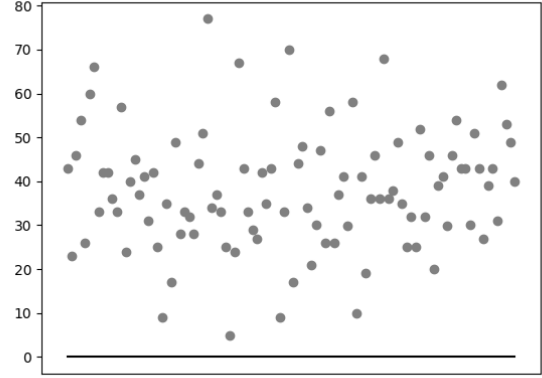
O efeito positivo esperado pelo *data augmentation* não se mostrou tão evidente quanto se esperava inicialmente. Porém, isto pode acontecer em razão dos valores dos hiperparâmetros

Figura 1.3: Resultados do treinamento e teste da CNN LeNet de acordo com a Abordagem 2.

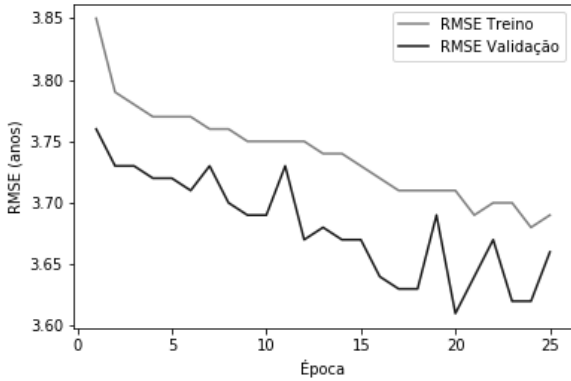
(a) RMSE de treinamento da arquitetura LeNet utilizando funções de ativação *ReLU*.



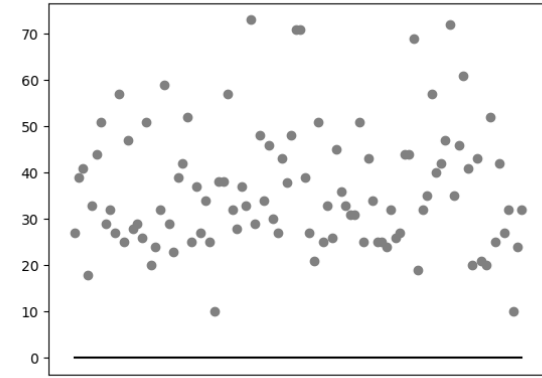
(b) Reta-0 LeNet *ReLU*.



(c) RMSE de treinamento da arquitetura LeNet utilizando funções de ativação *Leaky ReLU*.



(d) Reta-0 LeNet *Leaky ReLU*.



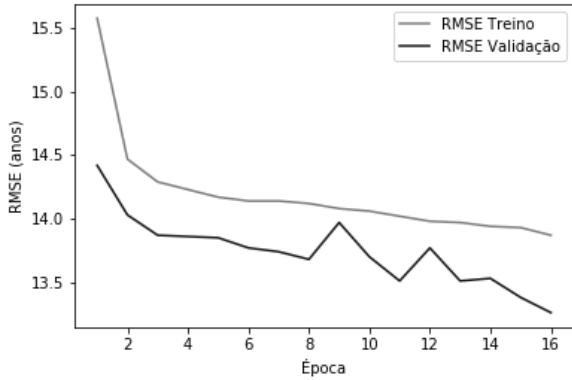
e da necessidade de melhor pré-processamento das imagens antes da apresentação às CNNs, o que motivou a realização da abordagem a seguir.

### 1.3 Abordagem 3: Introduzindo Equalização de Histograma

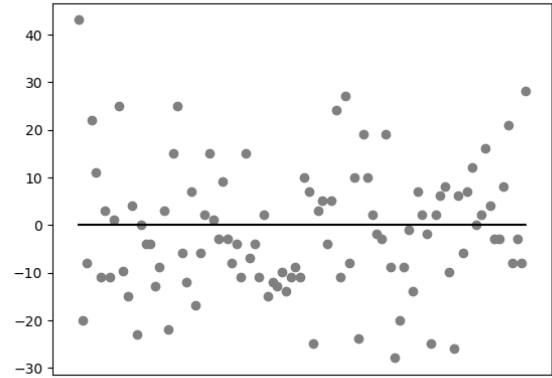
A terceira abordagem utilizou as imagens da base de dados normalizadas e técnicas de *data augmentation* previamente mencionadas. Considerando os resultados obtidos na abordagem anterior, introduziu-se o processo de equalização das imagens por histograma, que ajusta o

Figura 1.4: Resultados do treinamento e teste da CNN AlexNet de acordo com a Abordagem 2.

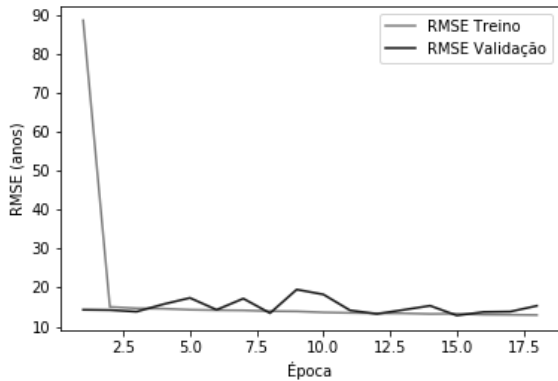
(a) Treinamento AlexNet *ReLU*.



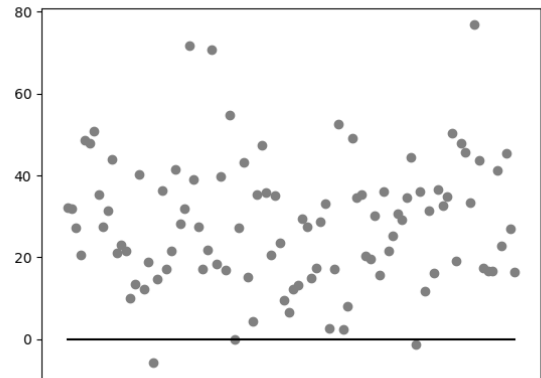
(b) Reta-0 AlexNet *ReLU*.



(c) Treinamento AlexNet *Leaky ReLU*.



(d) Reta-0 AlexNet *Leaky ReLU*.



contraste da imagem utilizando o histograma de cores. Este método aumenta o contraste global de imagens, especialmente quando os dados úteis da imagem são representados por cores próximas. Isto faz com que áreas de contraste menor ganhem mais contraste. No contexto da detecção de idade por meio da imagem da face de determinado indivíduo, a equalização por histograma reforça marcas de expressões e outras imperfeições (ACHARYA; RAY, 2005).

A partir desta abordagem foram obtidos os gráficos de treinamento e a reta-0 das redes LeNet e AlexNet, que estão nas Figuras 1.5 e 1.6, respectivamente.

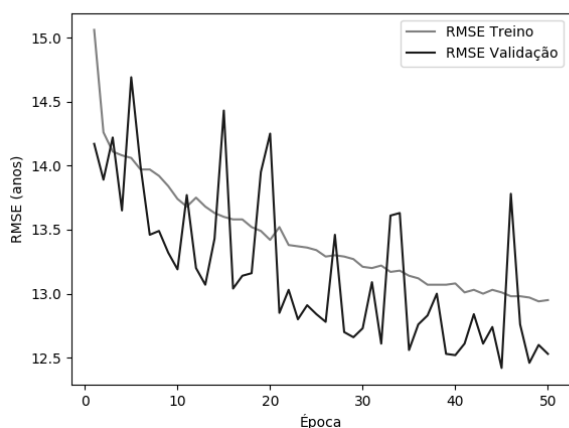
Obedecendo ao método de validação cruzada *holdout* previamente mencionado, os resultados desta abordagem encontram-se sintetizados na Tabela 1.3.

Tabela 1.3: Resultados do treino e teste dos modelos propostos na Abordagem 3.

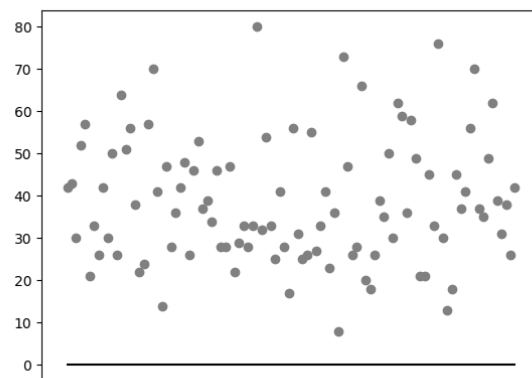
Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
LeNet	<i>ReLU</i>	46	38.66	41.20
LeNet	<i>Leaky ReLU</i>	38	38.26	40.85
AlexNet	<i>ReLU</i>	7	13.10	15.88
AlexNet	<i>Leaky ReLU</i>	18	35.25	38.04

Figura 1.5: Resultados do treinamento e teste da CNN LeNet de acordo com a Abordagem 3.

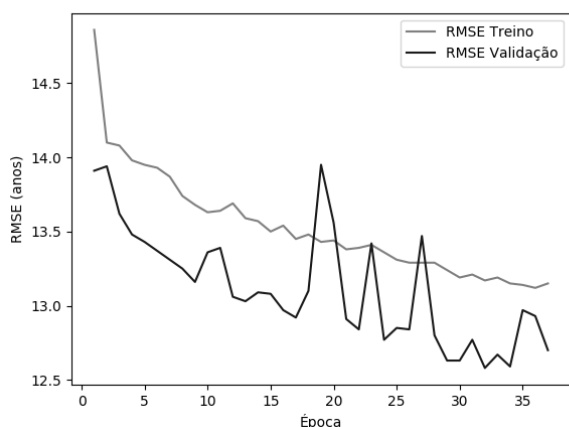
(a) RMSE de treinamento da arquitetura LeNet utilizando funções de ativação *ReLU*.



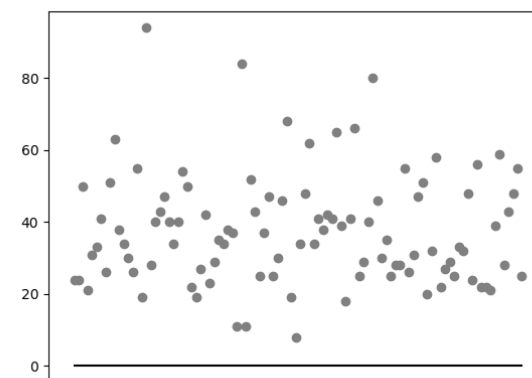
(b) Reta-0 LeNet *ReLU*.



(c) RMSE de treinamento da arquitetura LeNet utilizando funções de ativação *Leaky ReLU*.



(d) Reta-0 LeNet *Leaky ReLU*.





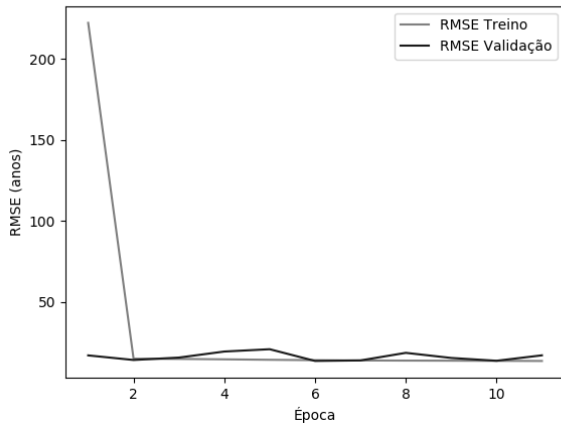
## 1.4 Abordagem 4: Utilizando MAE para o cálculo da perda

A análise dos gráficos de treinamento das redes anteriores levou à suposição de que a métrica utilizada para a atualização dos pesos RMSE estivesse trazendo instabilidade para o treinamento. Desta maneira, esta abordagem considera o treinamento da rede LeNet com *data augmentation*, imagens normalizadas e com equalização por histograma, utilizando MAE para cálculo da perda. Os gráficos do treinamento e reta-0 desta abordagem estão nas Figuras 1.7.

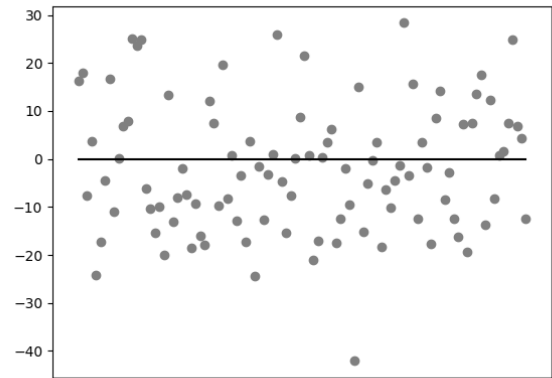
Obedecendo ao método de validação cruzada *holdout* previamente mencionado, os resultados

Figura 1.6: Resultados do treinamento e teste da CNN AlexNet de acordo com a Abordagem 3.

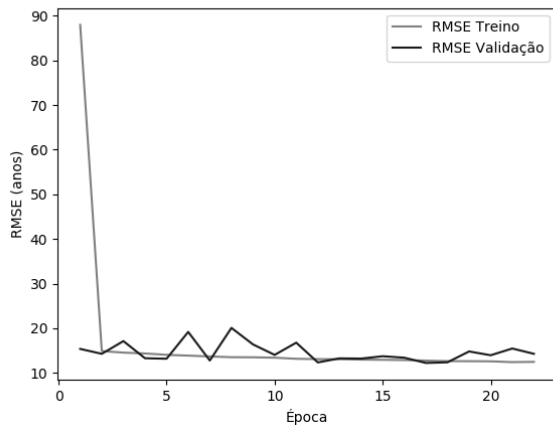
(a) Treinamento AlexNet *ReLU*.



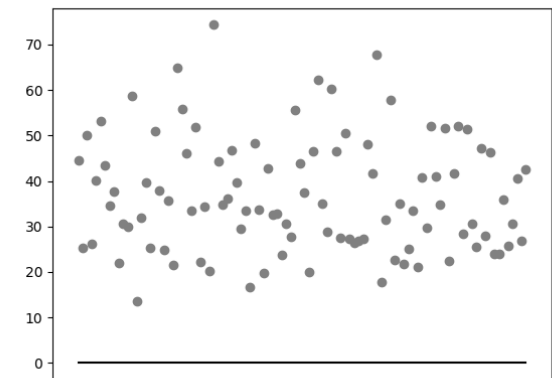
(b) Reta-0 AlexNet *ReLU*.



(c) Treinamento AlexNet *Leaky ReLU*.



(d) Reta-0 AlexNet *Leaky ReLU*.



desta abordagem encontram-se sintetizados na Tabela 1.4.

Tabela 1.4: Resultados do treino e teste dos modelos propostos na Abordagem 4.

Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
LeNet	<i>Leaky ReLU</i>	38	9.98	12.91

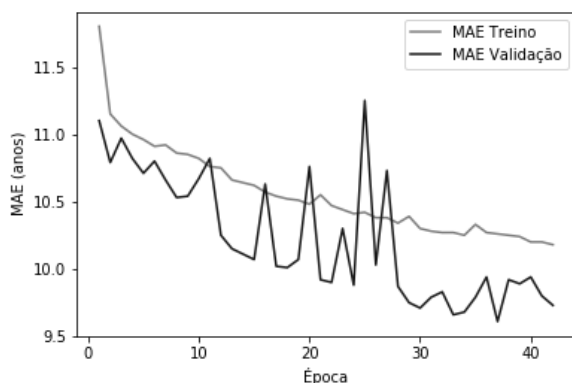
É possível notar que esta abordagem trouxe métricas de desempenho mais satisfatórias que as obtidas até então. Porém, as escolhas da rede, função de ativação e pré-processamento de imagens de entrada foram feitas a partir do palpite de que *data augmentation* e equalização por histograma trariam redes mais fortes. Assim, há a necessidade de verificar o desempenho de uma rede similar, porém sem estes processos.

## 1.5 Abordagem 5: Rede com melhor desempenho

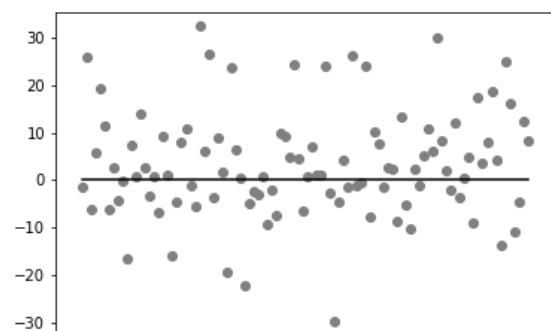
A quinta abordagem adotou a rede com melhor desempenho obtido até o momento. Neste caso, considerou-se a rede LeNet treinada com imagens da base de dados normalizadas, mas sem equalização de histograma de cores ou técnicas de *data augmentation*. Seguiu-se utilizando a métrica MAE para o cálculo da perda e da atualização dos pesos. Buscando garantir maior estabilidade nas métricas de desempenho durante o treinamento, aumentou-se o tamanho do batch para 128, haja vista a característica instável do treinamento mostrada nas abordagens

Figura 1.7: Resultados do treinamento e teste da CNN LeNet de acordo com a Abordagem 4.

(a) RMSE de treinamento da arquitetura LeNet utilizando funções de ativação *ReLU*.



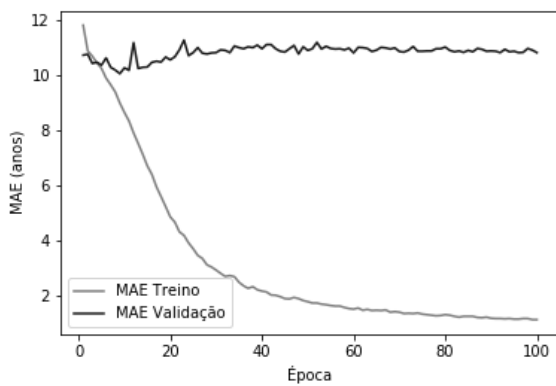
(b) Reta-0 LeNet *ReLU*.



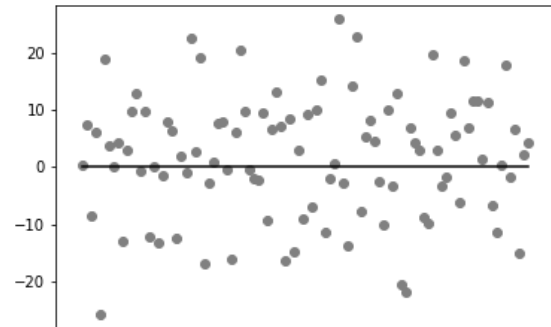
anteriores.

Figura 1.8: Resultados do treinamento e teste da CNN LeNet de acordo com a Abordagem 5.

(a) RMSE de treinamento da arquitetura LeNet utilizando funções de ativação *ReLU*.



(b) Reta-0 LeNet *ReLU*.



Obedecendo ao método de validação cruzada *holdout* previamente mencionado, os resultados desta abordagem encontram-se sintetizados na Tabela ??.

Tabela 1.5: Resultados do treino e teste dos modelos propostos na Abordagem 5.

Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
LeNet	<i>ReLU</i>	9	10.09	13.04

## 1.6 Abordagem x

- Mesmas redes - Normalização das imagens, equalização por histograma -> o que é - data augmentation -> mais técnicas de data augmentation

## 1.7 Abordagem x+1

Outras arquiteturas VGG com transfer learning 1. Retirar última camada (softmax) e adicionar leaky relu 2. Retirar duas últimas camadas (dense e softmax) e adicionar leaky relu

## Capítulo 2

# Considerações Finais

O objetivo deste trabalho consiste em elaborar estratégias inteligentes para estimação de idade de telespectadores de *Smart* TVs a partir de suas respectivas fotografias faciais. Para este fim, foram propostos, treinados e testados em caráter preliminar dois modelos de CNNs já bem estabelecidos na literatura, a LeNet e AlexNet, com dois perfis de hiperparâmetros cada um.

Com isto, observou-se uma melhora significativa na performance da AlexNet, enquanto o RMSE da LeNet não sofreu grandes mudanças. Quanto às saídas das redes, a LeNet exibiu valores positivos e negativos próximos de zero, e a AlexNet forneceu previsões que giravam em torno da média dos dados. Estes resultados são preliminares e certamente outros modelos e parâmetros serão investigados conforme previsto na metodologia e cronograma deste trabalho de conclusão de curso.

Nos próximos meses, os esforços estarão concentrados em pesquisar e adotar estratégias que possam minimizar os problemas identificados, como substituir as funções de ativação das camadas ocultas por outras variantes da *ReLU*, adotar métodos específicos de inicialização de pesos, normalização de *batch*, entre outros. Planeja-se também a proposição, o treinamento e teste de outras redes inspiradas em outros modelos canônicos.

O problema em questão é importante do ponto de vista prático para o desenvolvimento de diversas soluções de recomendação de conteúdo e controle parental em *Smart* TVs, auxiliando no desenvolvimento destas soluções tecnológicas. Considerando a formação de uma bacharelada em Engenharia de Computação, endereçar este problema permite a prática de diversos conceitos

vistos ao longo do curso, em especial relacionados às disciplinas de Inteligência Artificial, Redes Neurais, Processamento Digital de Imagens, *Machine Learning* e Sinais e Sistemas.

# Referências Bibliográficas

ACHARYA, T.; RAY, A. K. *Image processing: principles and applications*. [S.l.]: John Wiley & Sons, 2005.

CHOLLET, F. *Deep Learning with Python*. 1st. ed. Greenwich, CT, USA: Manning Publications Co., 2017. ISBN 1617294438, 9781617294433.