

Estimação Inteligente de Idade de Telespectadores para Aplicações de Sugestão de Conteúdo em Smart TVs

Trabalho de Conclusão de Curso II

por

Nicoli P. Araújo, Elloá B. Guedes

{npda.eng, ebgcosta}@uea.edu.br

do

Núcleo de Computação
Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Manaus – Amazonas – Brasil

Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados e Discussões

6. Considerações Finais

Contextualização e Motivação

- ***Smart TVs***
 - Conexão à internet
 - Transmissão de conteúdos advindos de outros dispositivos eletrônicos
- Pesquisa Nacional por Amostra de Domicílios 2015
 - *Smart TVs* representam 68,2% do total de televisores vendidos no primeiro semestre de 2017
- Aumento da popularidade causado por:
 - *Smart TVs versus* televisões convencionais
 - Encerramento da transmissão de sinal analógico da televisão aberta
 - Copa do Mundo 2018
 - Tecnologia 4K

Contextualização e Motivação

- ▶ Sistemas inteligentes de identificação de características:
 - ▶ Dificuldades práticas
 - ▶ Algumas *Smart TVs* possuem câmeras
 - ▶ Aplicações: Recomendação de conteúdo e controle parental

Objetivos

Objetivo Geral

Desenvolvimento de estratégias inteligentes, baseadas na utilização de técnicas de *Deep Learning*, para estimativa da idade de telespectadores a partir de fotografias faciais

Objetivos

Objetivo Geral

Desenvolvimento de estratégias inteligentes, baseadas na utilização de técnicas de *Deep Learning*, para estimativa da idade de telespectadores a partir de fotografias faciais

Objetivos Específicos

1. Referencial teórico sobre redes neurais convolucionais
2. Consolidar uma base de dados
3. Identificar tecnologias adequadas
4. Propor, treinar e testar diferentes estimadores de idade
5. Avaliar comparativamente os estimadores propostos

Justificativas

- Melhor experiência de provimento de conteúdo e de configurações personalizadas
- Prática e a proposição de soluções envolvendo *Machine Learning*
- Laboratório de Sistemas Inteligentes (LSI)

Metodologia

- Fundamentação teórica sobre Machine Learning
- Fluxo de atividades de *machine learning* (Marsland, 2015)
 - Consolidar uma base de dados
 - Propor modelos aplicáveis
 - Treinamento
 - Teste
 - Comparação de métricas de desempenho
- Escrita da proposta e do projeto final do trabalho de conclusão de curso

Cronograma

Tabela 1: Cronograma de atividades

	2018										
	02	03	04	05	06	07	08	09	10	11	12
Escrita da Proposta	X	X	X	X	X						
Fundamentação Teórica sobre ML	X	X	X	X							
Consolidação da Base de Dados		X	X								
Proposição de Modelos de CNNs				X	X	X	X	X			
Defesa da Proposta					X						
Escrita do Trabalho Final						X	X	X	X	X	X
Treinamento das CNNs					X	X	X	X	X	X	
Teste das CNNs					X	X	X	X	X	X	X
Comparação de Métricas de Desempenho						X	X	X	X	X	X
Defesa do Trabalho Final											X

Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados e Discussões

6. Considerações Finais

Redes Neurais Artificiais

McCulloch e Pitts, 1943

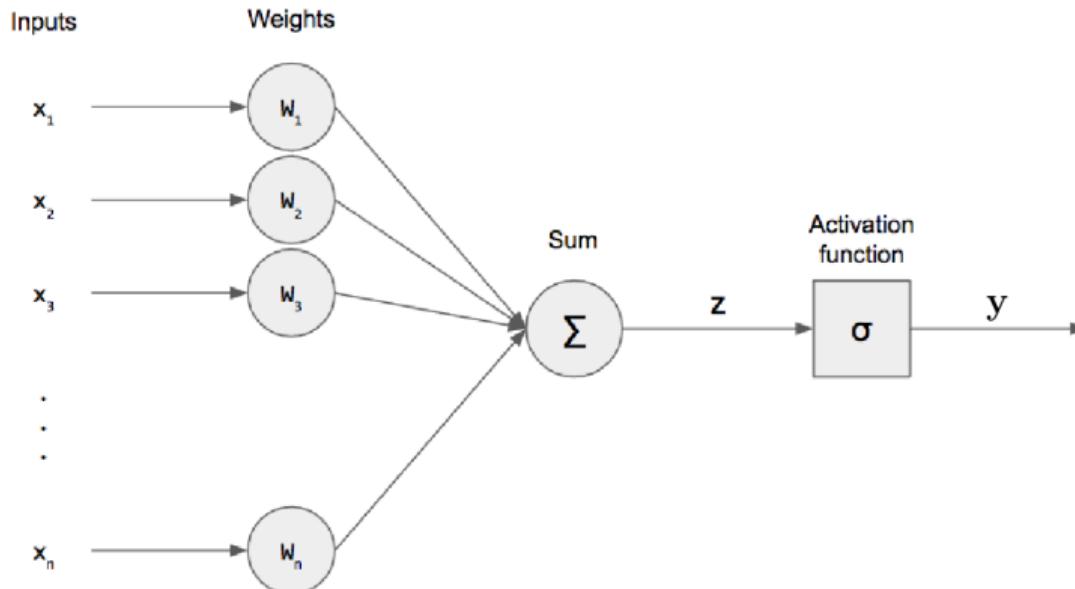


Figura 1: Representação de um neurônio artificial

Redes Neurais Convolucionais

- Topologia bem definida e estrutura em grid
- Destaca-se no reconhecimento de padrões em dados de alta dimensionalidade

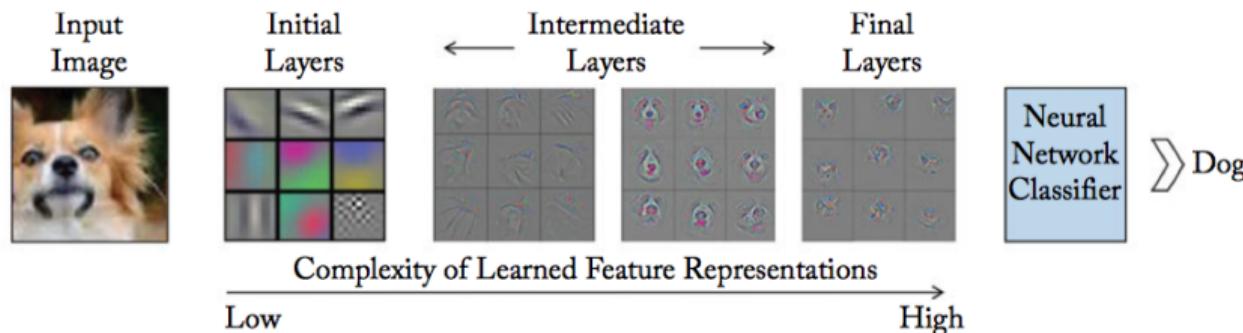


Figura 2: Papel das camadas convolucionais e *feature maps* nas CNNs.

Modelos Canônicos de Redes Neurais Convolucionais

- Arquiteturas que trouxeram contribuições importantes
- Comuns ainda hoje no cenário de DL
- LeNet (1998)
- AlexNet (2012)
- VGG (2014)
- Inception (2014)
- ResNet (2015)
- *Transfer Learning*: Aproveitamento de parâmetros treinados

Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados e Discussões

6. Considerações Finais

Trabalhos Relacionados

‣ DEX (Rothe,2015)

- Problema de classificação: idade entre 0 e 100
- Datasets: *IMDb*, *Wikipedia* e *Looking At People*
- Ensemble de 20 VGG-16, *Transfer Learning*
- MAE (*Mean Average Error*) de 3.221 na fase de testes

Trabalhos Relacionados

‣ **DEX** (Rothe,2015)

- Problema de classificação: idade entre 0 e 100
- Datasets: *IMDb*, *Wikipedia* e *Looking At People*
- Ensemble de 20 VGG-16, *Transfer Learning*
- MAE (*Mean Average Error*) de 3.221 na fase de testes

‣ **AgeNet** (Liu et al., 2015)

- Ensemble de modelos regressores e classificadores com diferentes métodos de normalização de face
- CNNs inspiradas na *GoogLeNet*
- MAE de 3.3345

Trabalhos Relacionados

➢ Levi e Hassner, 2015:

- Idade e gênero
- Problema de classificação: 8 classes de idade
- Três camadas convolucionais com *dropout* e funções de ativação *ReLU*, seguidas por três camadas totalmente conectadas
- Acurácia de $84.7\% \pm 2.2$

Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados e Discussões

6. Considerações Finais

Tarefa de Aprendizado

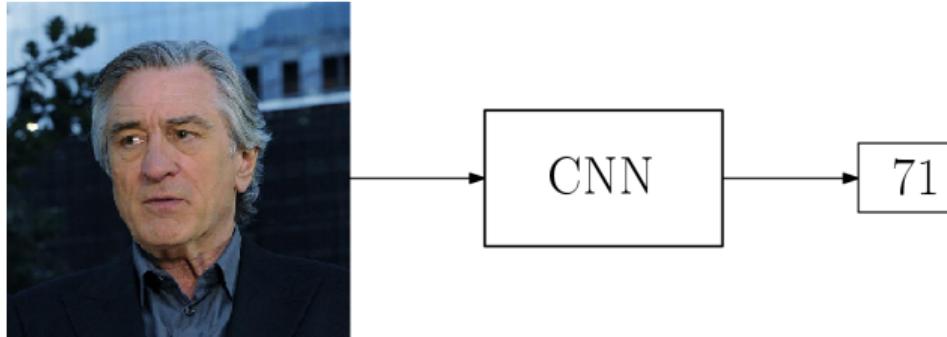
- Tarefa de Regressão

- Entrada:

- Imagem em cores RGB de dimensões 224×224 pixels contendo uma face humana centralizada

- Saída:

- Estimativa de idade, em anos, da pessoa correspondente



Tarefa de Aprendizado

- Métricas de desempenho:

- *Root Mean Squared Error (RMSE)*

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2}. \quad (1)$$

- *Mean Absolute Error (MAE)*

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}| \quad (2)$$

Conjunto de Dados

- Base de dados experimentais IMDb
 - 452.132 exemplos
 - 20.284 dos atores mais populares listados no site IMDb
 - Organizada por Rothe et al., 2015
 - Imagens e meta-dados

Conjunto de Dados

Meta-dado	Valor
ID Celebridade	16349
Nome	Robert De Niro
Endereço da imagem	imdb/34/nm0000134_rm334009 0368_1943-8-17_2011.jpg
Pontuação da Face	5.21396
Pontuação da Segunda Face	NaN
Localização da Face	(663.65, 992.475, 590.134, 918.959)
Data de Nascimento	1943 – 08 – 17
Ano da Foto	2011
Gênero	Masculino

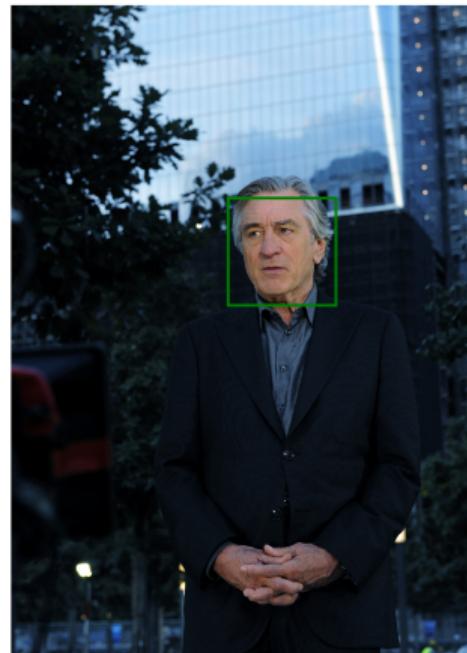


Figura 3: Exemplo de imagem do conjunto de dados contendo apenas um rosto.

Conjunto de Dados

Meta-dado	Valor
ID Celebridade	16349
Nome	Robert De Niro
Endereço da imagem	imdb/34/nm0000134_rm14800 44288_1943-8-17_2012.jpg
Pontuação da Face	5.51656
Pontuação da Segunda Face	4.55379
Localização da Face	(1392.72, 1614.18, 225.55, 447.003)
Data de Nascimento	1943 – 08 – 17
Ano da Foto	2012
Gênero	Masculino



Figura 4: Exemplo de imagem do conjunto de dados contendo mais de um rosto com a classificação errônea.

Limpeza e Pré-processamento dos dados

- Versão original: 267 GB
- Faces recortadas: 7,1 GB
- Cálculo do atributo alvo: **Idade**
- Itens descartados:
 - Idade e gênero apresentando valores nulos, inválidos ou negativos
 - Múltiplos exemplos referentes à mesma pessoa com a mesma idade
 - Idade maior que 100 anos ou não compatível com os dados da celebridade referida nos meta-dados
 - Ausência de rosto
 - Presença de mais de uma face na imagem

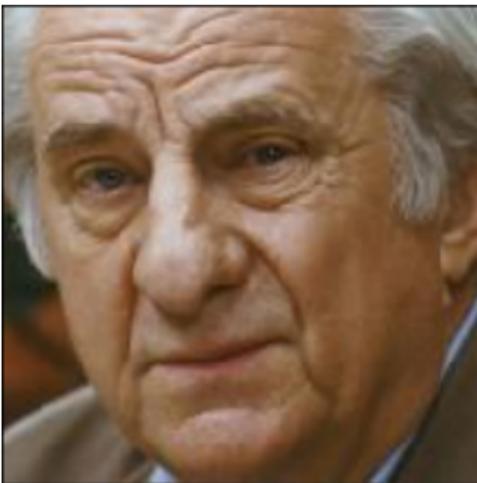
Limpeza e Pré-processamento dos dados

- Padronização das dimensões das imagens
 - 224×224 pixels
 - RGB
- Descarte de meta-dados irrelevantes para a tarefa de aprendizado

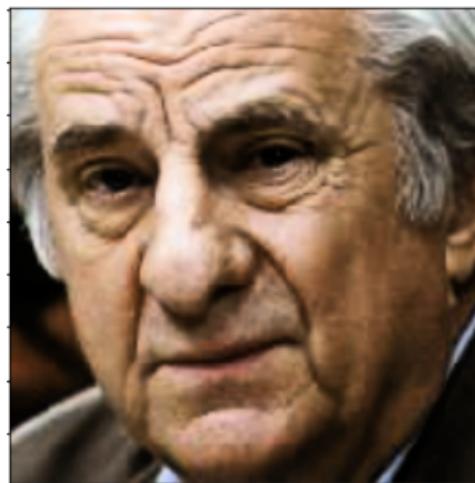
Limpeza e Pré-processamento dos dados

‣ Equalização por histograma

(a) Imagem sem equalização de histograma.



(b) Imagem após equalização por histograma.



Limpeza e Pré-processamento dos dados

- Conjunto de dados consolidado:
 - 47.950 exemplos
 - 14.607 celebridades
 - 1,2GB em disco

- Divisão obedecendo o método *Holdout*
 - Treinamento–Validação–Teste
 - 70% – 10% – 20%
 - 33.565 – 4.795 – 9.590

Modelos de CNN Considerados

▣ Arquiteturas

- LeNet
- AlexNet
- VGG-16
- SqueezeNet

Modelos de CNN Considerados

- Funções de ativação: *ReLU* ou *Leaky ReLU*
- Camadas de saída com apenas um neurônio
- Número de épocas obtido de maneira experimental
- Método de otimização do gradiente descendente *Adam*
- *batch size* igual a 64
- Taxa de aprendizado de 0.001

Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados e Discussões

6. Considerações Finais

Resultados e Discussão

- Estimação de idade a partir de uma imagem de face
- Arquiteturas
 - LeNet
 - AlexNet
 - VGG-16
 - SqueezeNet
- **9 Abordagens**

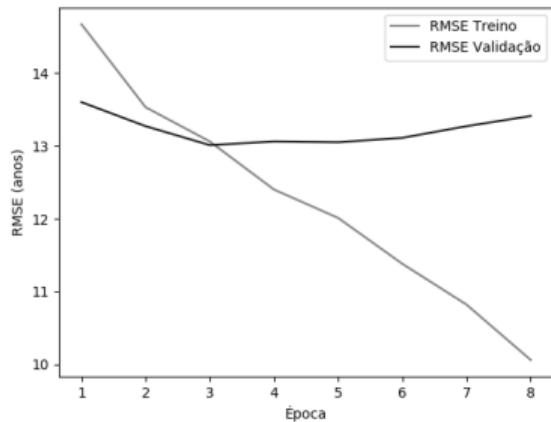
Abordagem 1: LeNet e AlexNet com Imagens Normalizadas

- Entradas: Imagens normalizadas
- Redes: LeNet e AlexNet
- Funções de ativação: *ReLU* e *Leaky ReLU*

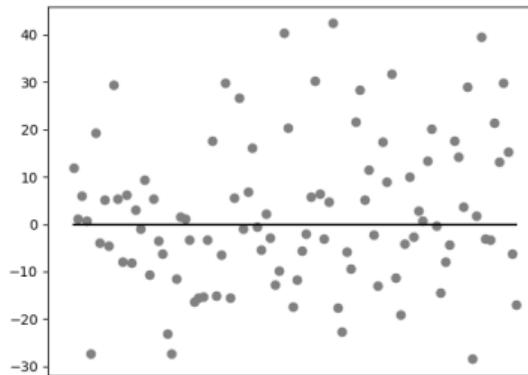
Abordagem 1: LeNet e AlexNet com Imagens Normalizadas

Figura 6: Resultados do treinamento e teste da CNN LeNet *ReLU*

(a) RMSE de treinamento



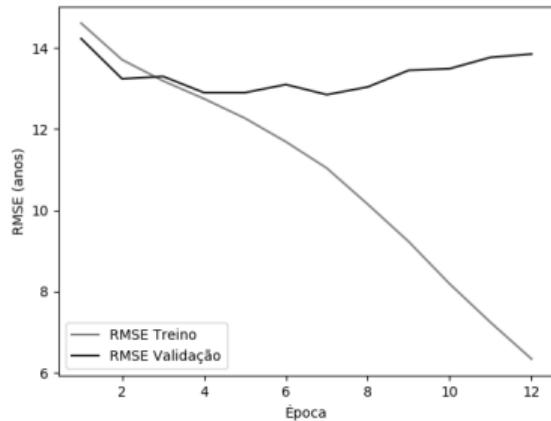
(b) Reta-0 do conjunto de teste



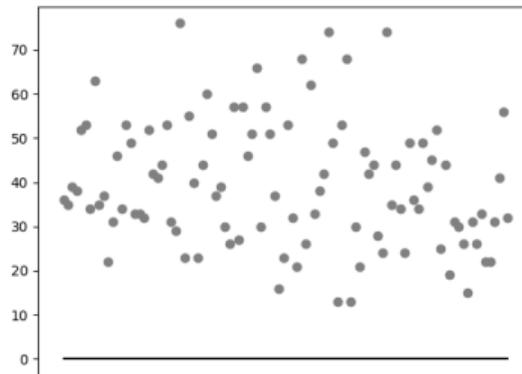
Abordagem 1: LeNet e AlexNet com Imagens Normalizadas

Figura 8: Resultados do treinamento e teste da CNN LeNet *Leaky ReLU*

(a) RMSE de treinamento



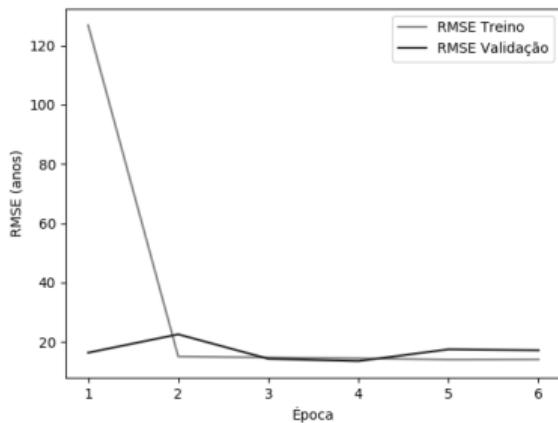
(b) Reta-0 do conjunto de teste



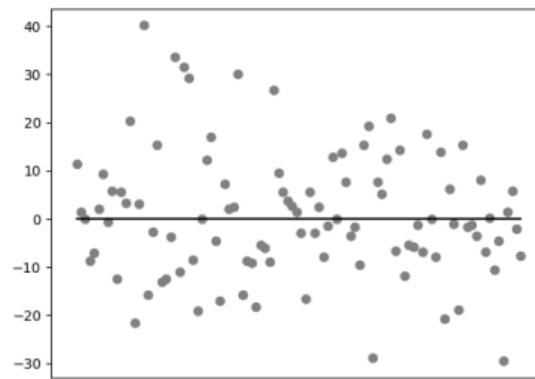
Abordagem 1: LeNet e AlexNet com Imagens Normalizadas

Figura 10: Resultados do treinamento e teste da CNN AlexNet *ReLU* de acordo com a Abordagem 1

(a) RMSE Treinamento



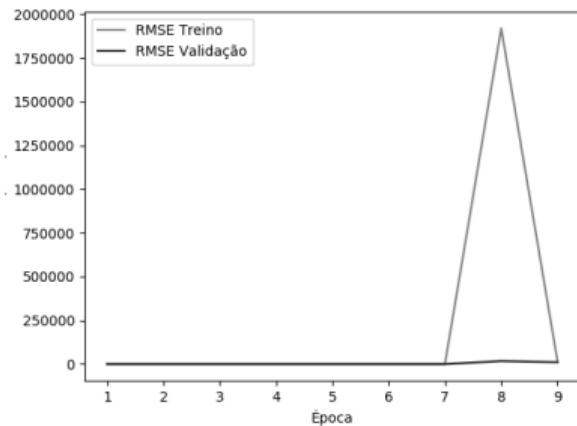
(b) Reta-0



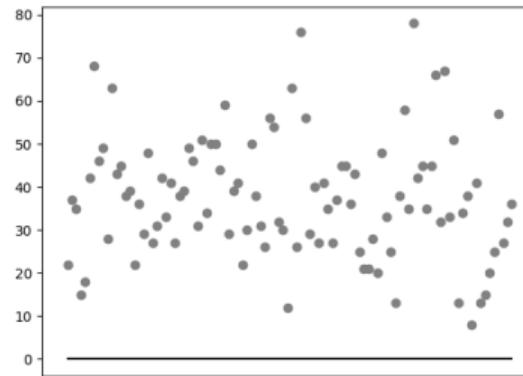
Abordagem 1: LeNet e AlexNet com Imagens Normalizadas

Figura 12: Resultados do treinamento e teste da CNN AlexNet *Leaky ReLU* de acordo com a Abordagem 1

(a) RMSE Treinamento



(b) Reta-0



Abordagem 1: LeNet e AlexNet com Imagens Normalizadas

Tabela 2: Resultados do treino e teste dos modelos propostos na Abordagem 1

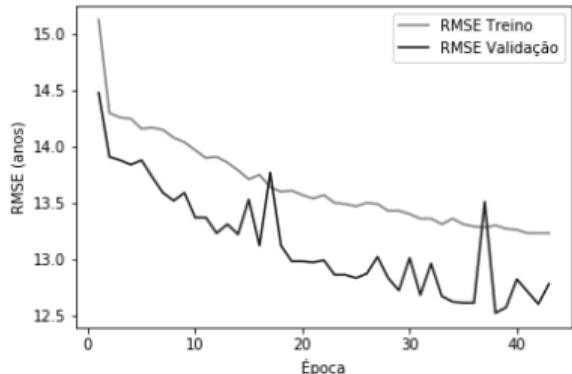
Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
LeNet	<i>ReLU</i>	4	10.53	13.55
LeNet	<i>Leaky ReLU</i>	8	38.33	40.82
AlexNet	<i>ReLU</i>	5	11.03	13.76
AlexNet	<i>Leaky ReLU</i>	5	39.27	41.97

- Entradas: Imagens normalizadas, ***data augmentation***
 - Rotação, zoom, inversão horizontal e translação
- Redes: LeNet e AlexNet
- Funções de ativação: *ReLU* e *Leaky ReLU*

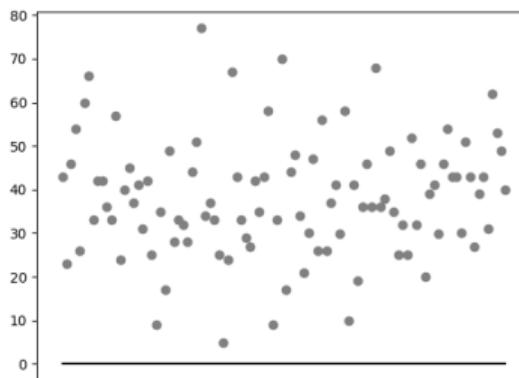
Abordagem 2: Introduzindo *Data Augmentation*

Figura 14: Resultados do treinamento e teste da CNN LeNet ReLU

(a) RMSE de treinamento



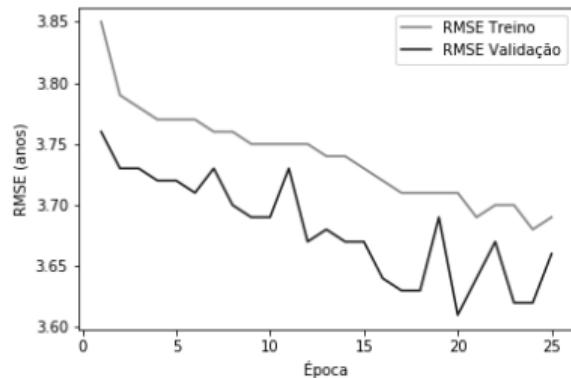
(b) Reta-0 do conjunto de teste



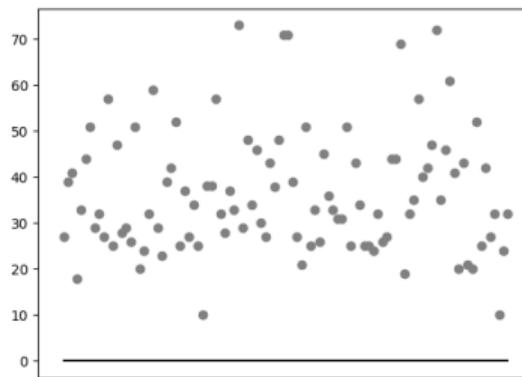
Abordagem 2: Introduzindo *Data Augmentation*

Figura 16: Resultados do treinamento e teste da CNN LeNet *Leaky ReLU*

(a) RMSE de treinamento



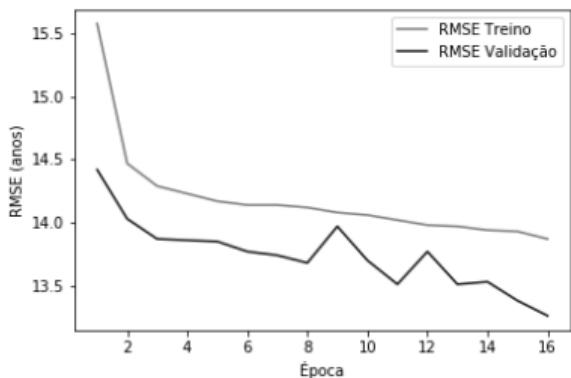
(b) Reta-0 do conjunto de teste



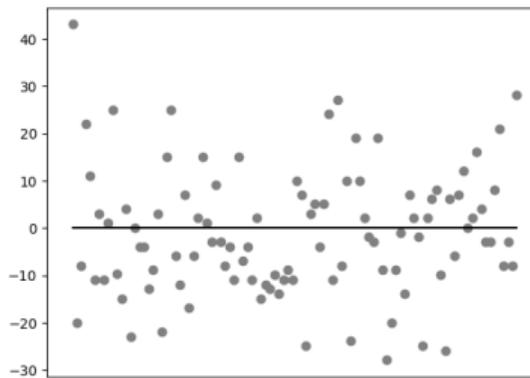
Abordagem 2: Introduzindo *Data Augmentation*

Figura 18: Resultados do treinamento e teste da CNN AlexNet *ReLU* de acordo com a Abordagem 2

(a) RMSE Treinamento



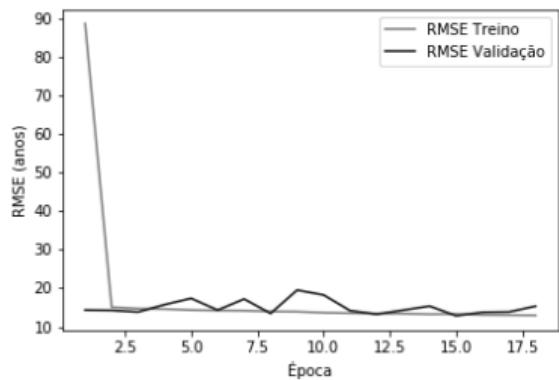
(b) Reta-0



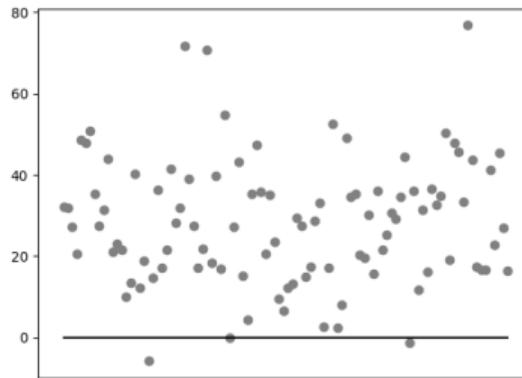
Abordagem 2: Introduzindo *Data Augmentation*

Figura 20: Resultados do treinamento e teste da CNN AlexNet *Leaky ReLU* de acordo com a Abordagem 2

(a) RMSE Treinamento



(b) Reta-0



Abordagem 2: Introduzindo *Data Augmentation*

Tabela 3: Resultados do treino e teste dos modelos propostos na Abordagem 1

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

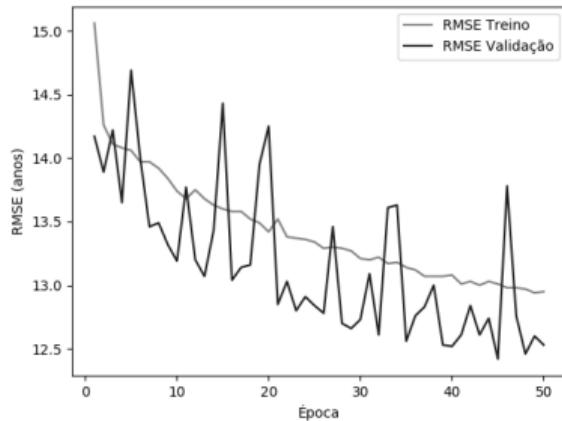
Abordagem 3

- Entradas: Imagens normalizadas, **equalização por histograma**, *data augmentation*
 - ▶ Rotação, zoom, inversão horizontal e translação
- Redes: LeNet e AlexNet
- Funções de ativação: *ReLU* e *Leaky ReLU*

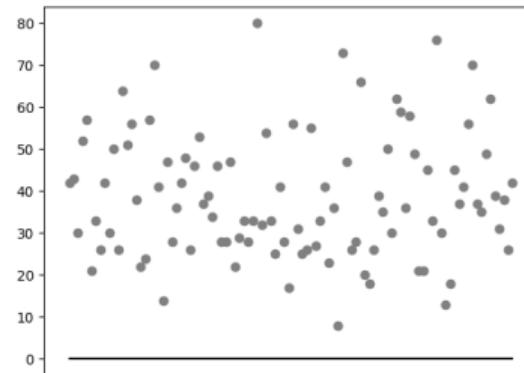
Abordagem 3: Introduzindo Equalização de Histograma

Figura 22: Resultados do treinamento e teste da CNN LeNet *ReLU*

(a) RMSE de treinamento



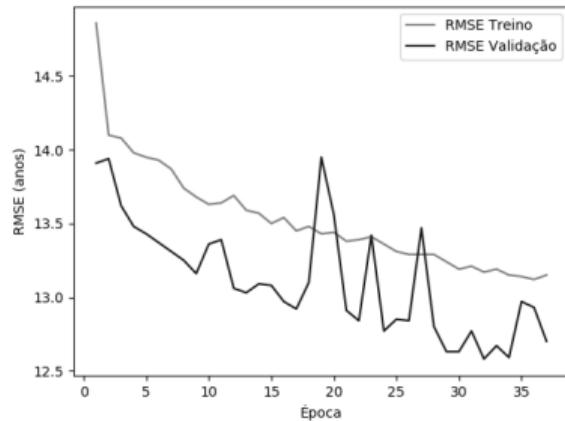
(b) Reta-0 do conjunto de teste



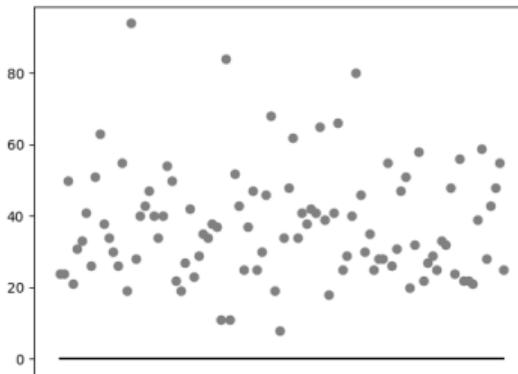
Abordagem 3: Introduzindo Equalização de Histograma

Figura 24: Resultados do treinamento e teste da CNN LeNet Leaky ReLU

(a) RMSE de treinamento



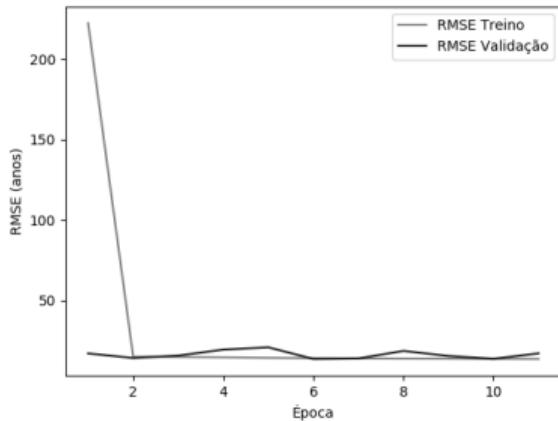
(b) Reta-0 do conjunto de teste



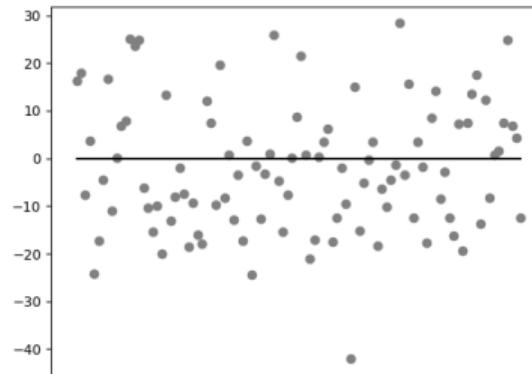
Abordagem 3: Introduzindo Equalização de Histograma

Figura 26: Resultados do treinamento e teste da CNN AlexNet *ReLU*

(a) RMSE Treinamento



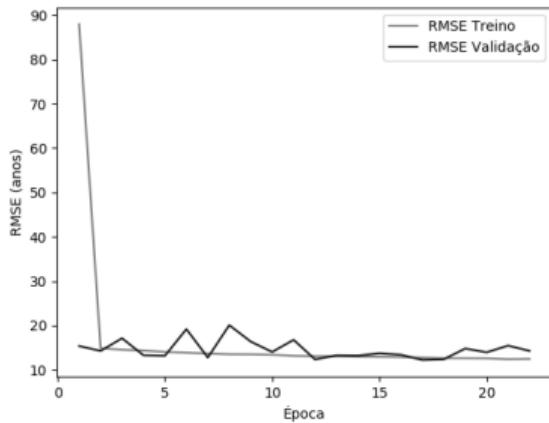
(b) Reta-0



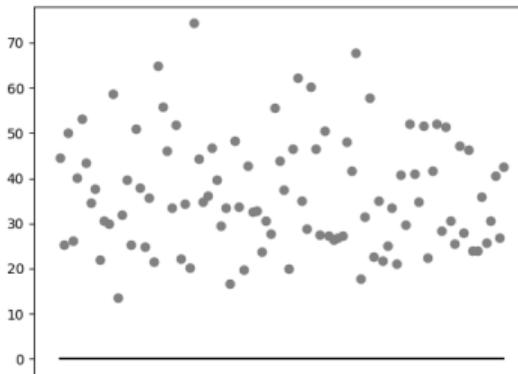
Abordagem 3: Introduzindo Equalização de Histograma

Figura 28: Resultados do treinamento e teste da CNN AlexNet *Leaky ReLU*

(a) RMSE Treinamento



(b) Reta-0



Abordagem 3: Introduzindo Equalização de Histograma

Tabela 4: Resultados do treino e teste dos modelos propostos na Abordagem 1

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

Abordagem 4

- Rede com melhor desempenho verificado até então
- **LeNet *ReLU* da Abordagem 1**

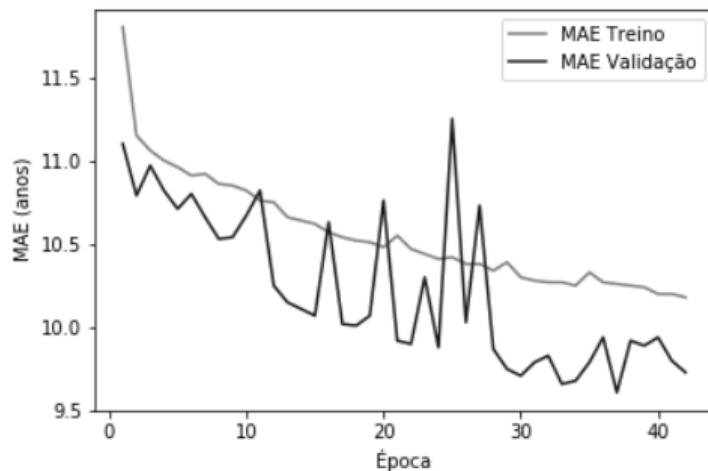
- Entradas: Imagens normalizadas, equalização por histograma, *data augmentation*
- Rede: LeNet
- Funções de ativação: *ReLU*

- Adoção do MAE para cálculo da perda
- Tamanho do batch alterado para 128 imagens

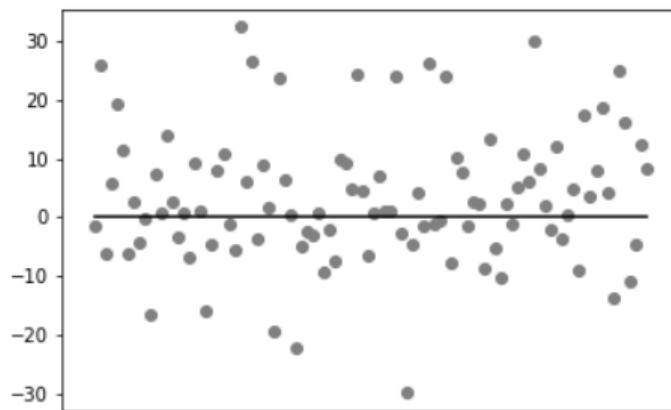
Abordagem 4: Utilizando MAE para o Cálculo da Perda

Figura 30: Resultados do treinamento e teste da CNN LeNet de acordo com a Abordagem 4

(a) MAE de treinamento



(b) Reta-0



Abordagem 4: Utilizando MAE para o Cálculo da Perda

Tabela 5: Resultados do treino e teste do modelo proposto na Abordagem 4

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

Abordagem 5

- Rede com melhor desempenho verificado até então
- **LeNet *ReLU* da Abordagem 1**

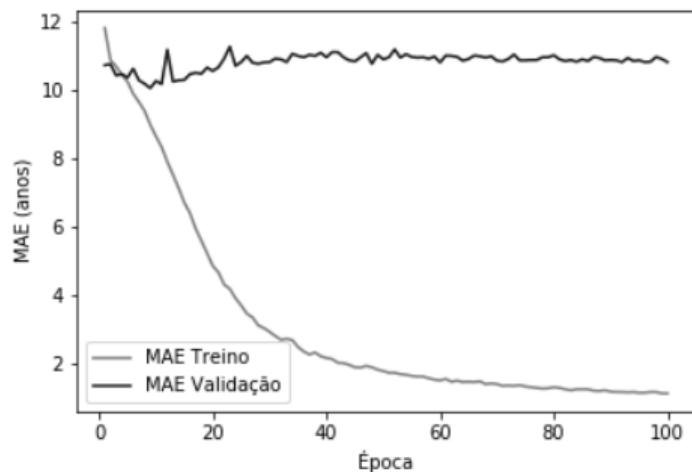
- Entradas: **Imagens normalizadas**
- Rede: LeNet
- Funções de ativação: *ReLU*

- Cálculo da perda considerando o MAE
- Tamanho do batch: 128 imagens

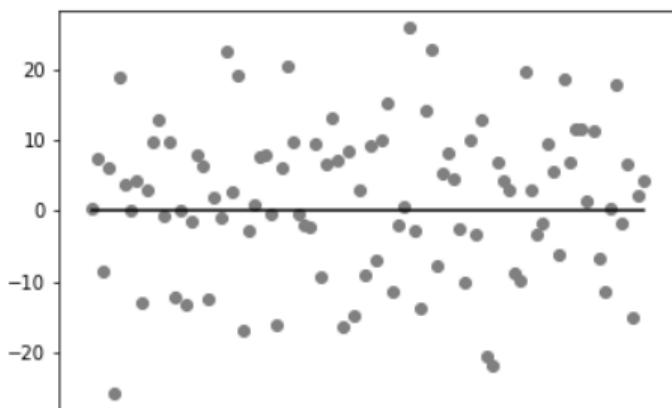
Abordagem 5: LeNet Apenas com Normalização da Entrada

Figura 32: Resultados do treinamento e teste da CNN LeNet de acordo com a Abordagem 5

(a) MAE de treinamento



(b) Reta-0



Abordagem 5: LeNet Apenas com Normalização da Entrada

Tabela 6: Resultados do treino e teste do modelo proposto na Abordagem 5

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

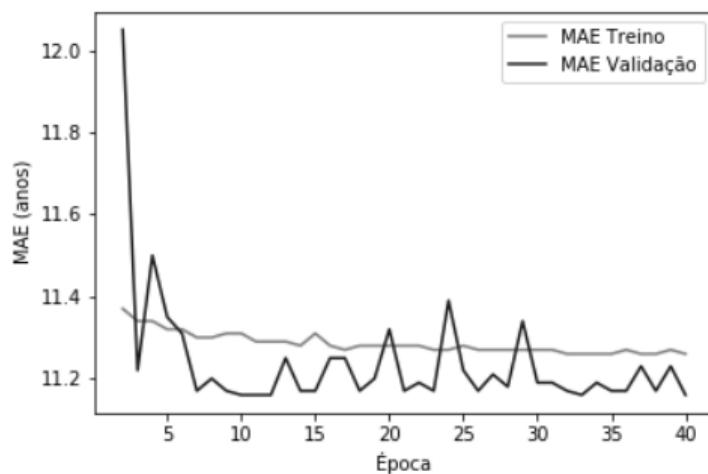
Abordagem 6

- Entradas: Imagens normalizadas
- Rede: **VGG-16**
- Funções de ativação: *ReLU*
- Cálculo da perda considerando o MAE
- **Tamanho do batch: 64 imagens**

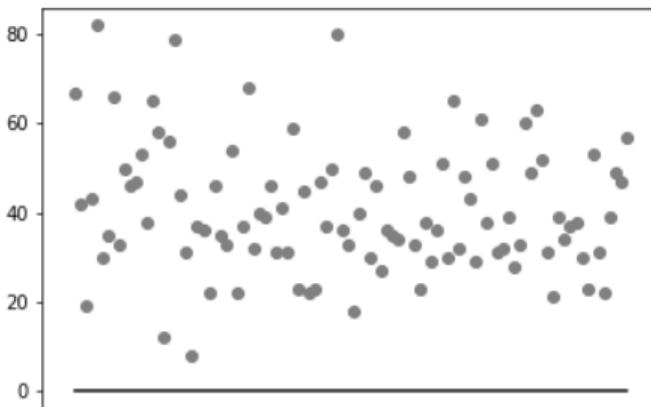
Abordagem 6: VGG-16 e Dados Normalizados

Figura 34: Resultados do treinamento e teste da CNN VGG-16 de acordo com a Abordagem 6

(a) MAE de treinamento da arquitetura VGG-16 utilizando funções de ativação *ReLU*



(b) Reta-0 VGG-16 *ReLU*



Abordagem 6: VGG-16 e Dados Normalizados

Tabela 7: Resultados do treino e teste dos modelos propostos na Abordagem 6

Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
VGG-16	<i>ReLU</i>	11	40.90	38.35

Abordagem 6: VGG-16 e Dados Normalizados

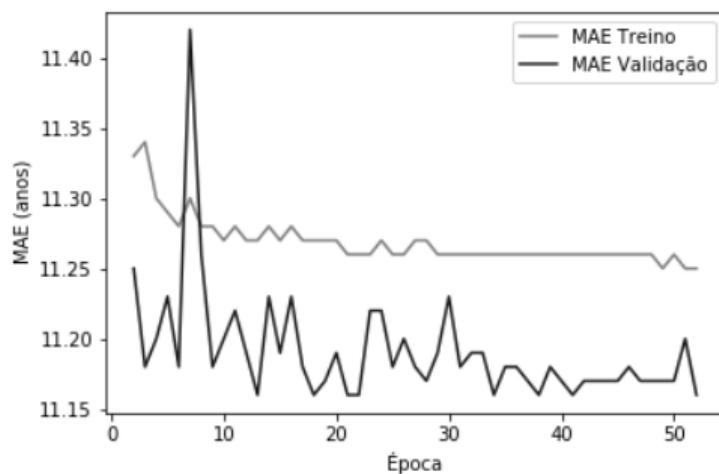
- *Dying ReLU problem*
- Saídas iguais a 37.05 para todas as imagens dadas como entrada

- Entradas: Imagens normalizadas, **equalização por histograma, data augmentation**
- Rede: VGG-16
- Funções de ativação: *ReLU*
- Cálculo da perda considerando o MAE
- Tamanho do batch: 64 imagens

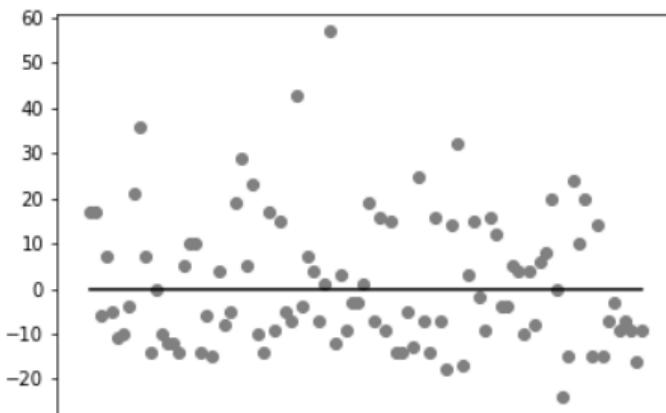
Abordagem 7: VGG-16 com *Data Augmentation* e Equalização de Histograma

Figura 36: Resultados do treinamento e teste da CNN VGG-16 de acordo com a Abordagem 7

(a) MAE de treinamento da arquitetura VGG-16 utilizando funções de ativação *ReLU*



(b) Reta-0 VGG-16 *ReLU*



Abordagem 7: VGG-16 com *Data Augmentation* e Equalização de Histograma

Tabela 8: Resultados do treino e teste dos modelos propostos na Abordagem 7

Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
VGG-16	<i>ReLU</i>	23	40.99	38.39

Abordagem 7: VGG-16 com *Data Augmentation* e Equalização de Histograma

- *Dying ReLU problem*
- Saídas iguais a 37.02 para todas as imagens dadas como entrada

Abordagem 8

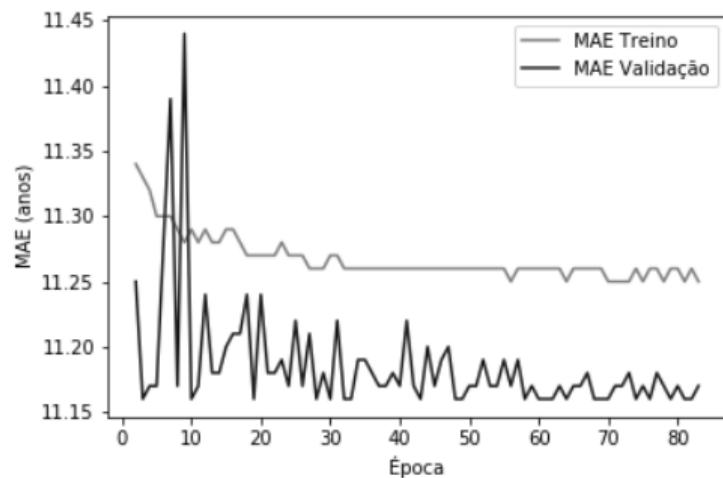
- Entradas: Imagens normalizadas, equalização por histograma, *data augmentation*
- Rede: VGG-16
- Função de ativação na camada de saída: **Leaky ReLU**
- **Transfer Learning: ImageNet**

- Cálculo da perda considerando o MAE
- Tamanho do batch: 64 imagens

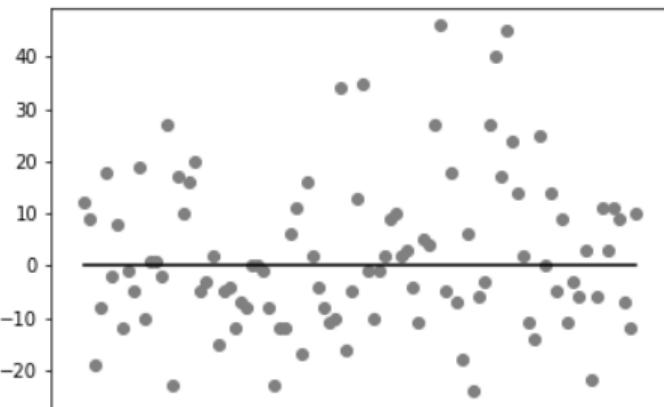
Abordagem 8: VGG-16 com *data augmentation*, Transfer Learning e Leaky ReLU

Figura 38: Resultados do treinamento e teste da CNN VGG-16 de acordo com a Abordagem 8

(a) MAE de treinamento da arquitetura VGG-16 utilizando funções de ativação *Leaky ReLU*



(b) Reta-0 VGG-16 *Leaky ReLU*



Abordagem 8: VGG-16 com *data augmentation*, *Transfer Learning* e *Leaky ReLU*

Tabela 9: Resultados do treino e teste dos modelos propostos na Abordagem 8

Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
VGG-16	<i>Leaky ReLU</i>	83	11.02	14.24

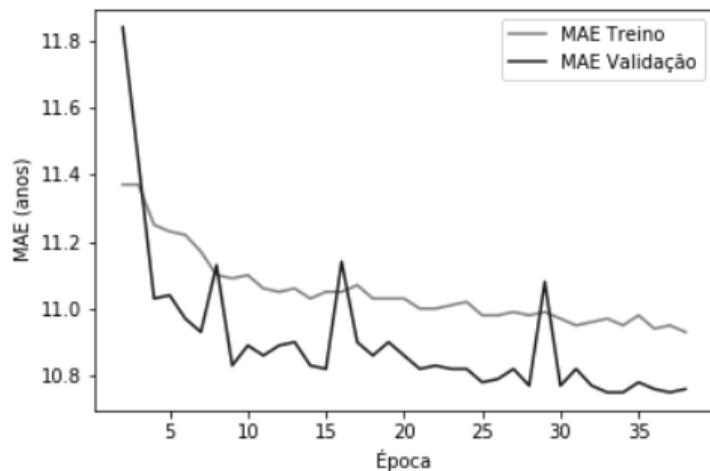
- *Dying ReLU problem*
- Saídas iguais a 36.99 para todas as imagens dadas como entrada

- Entradas: Imagens normalizadas, equalização por histograma, *data augmentation*
- Rede: **SqueezeNet**
- Função de ativação na camada de saída: *ReLU*
- Cálculo da perda considerando o MAE
- Tamanho do batch: 64 imagens

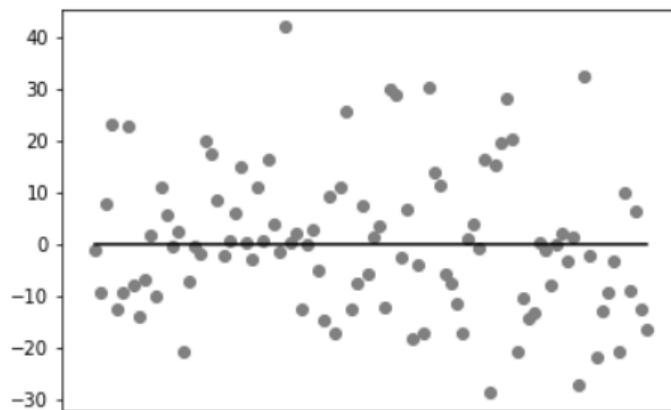
Abordagem 9: SqueezeNet

Figura 40: Resultados do treinamento e teste da CNN SqueezeNet de acordo com a Abordagem 9

(a) MAE de treinamento



(b) Reta-0



Abordagem 9: SqueezeNet

Tabela 10: Resultado do teste da SqueezeNet proposta na Abordagem 9

Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
SqueezeNet	<i>ReLU</i>	38	10.72	13.84

Sumarizando os Resultados

Tabela 11: Sumário dos resultados obtidos de todas as abordagens conduzidas.

Abordagem	Rede	Função de ativação	Épocas	MAE Teste	RMSE Teste
1	LeNet	<i>ReLU</i>	4	10.53	13.55
1	LeNet	<i>Leaky ReLU</i>	8	38.33	40.82
1	AlexNet	<i>ReLU</i>	5	11.03	13.76
1	AlexNet	<i>Leaky ReLU</i>	5	39.27	41.97
2	LeNet	<i>ReLU</i>	39	37.85	40.27
2	LeNet	<i>Leaky ReLU</i>	21	38.50	41.06
2	AlexNet	<i>ReLU</i>	16	11.59	14.59
2	AlexNet	<i>Leaky ReLU</i>	16	28.06	31.81
3	LeNet	<i>ReLU</i>	46	38.66	41.20
3	LeNet	<i>Leaky ReLU</i>	38	38.26	40.85
3	AlexNet	<i>ReLU</i>	7	13.10	15.88
3	AlexNet	<i>Leaky ReLU</i>	18	35.25	38.04
4	LeNet	<i>Leaky ReLU</i>	38	9.98	12.91
5	LeNet	<i>ReLU</i>	9	10.09	13.04
6	VGG-16	<i>ReLU</i>	11	40.90	38.35
7	VGG-16	<i>ReLU</i>	23	40.99	38.39
8	VGG-16	<i>Leaky ReLU</i>	83	11.02	14.24
9	SqueezeNet	<i>ReLU</i>	38	10.72	13.84

Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados e Discussões

6. Considerações Finais

Considerações Finais

➢ Resultados Preliminares

- Elaborar estratégias inteligentes para estimação de idade de telespectadores de *Smart TVs* a partir de suas respectivas fotografias faciais
- Modelos de CNNs considerados: LeNet, AlexNet, VGG-16, SqueezeNet
- 9 Abordagens

Considerações Finais

- Não foi possível constatar melhora progressiva
- Trabalho conduzido conforme práticas de natureza heurística sugeridas por diversos autores da literatura
- Melhor rede: **LeNet sujeita à abordagem 4**
 - Entradas: normalização, *data augmentation*, equalização por histograma
 - Função de ativação *ReLU*
 - MAE de 9.98 anos

Considerações Finais

- Condução assertiva da metodologia
- Prática de métodos e tecnologias em **Deep Learning**
- Patamar para comparação

Considerações Finais

Trabalhos futuros

- Tarefa de classificação
- Avaliação do impacto da utilização de técnicas de *transfer learning*
- Análise mais detalhada dos grupos de idade com maior e menor índice de erro
- Auxílio no desenvolvimento de soluções de recomendação de conteúdo e controle parental em *Smart TVs*

Estimação Inteligente de Idade de Telespectadores para Aplicações de Sugestão de Conteúdo em Smart TVs

Trabalho de Conclusão de Curso II

por

Nicoli P. Araújo, Elloá B. Guedes

{npda.eng, ebgcosta}@uea.edu.br

do

Núcleo de Computação
Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Manaus – Amazonas – Brasil