

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

**Trabalho de Conclusão de Curso I**

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 Parciais

6. Considerações Parciais

# Contextualização e Motivação

## ► Smart TVs

- Conexão à internet
- Transmissão de conteúdos advindos de outros dispositivos eletrônicos

## ► PNAD 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

## ➢ Recomendação de Conteúdo:

- Dificuldades práticas
- Algumas *Smart TVs* possuem câmeras
- Sistemas inteligentes de identificação de características
- Recomendação de conteúdo
- 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
  - Controle Parental
- 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
  - Proposição de diferentes modelos de redes neurais convolucionais
  - Treinamento das redes neurais convolucionais
  - Teste das redes
  - 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
<b>Escrita da Proposta</b>	X	X	X	X	X						
<b>Fundamentação Teórica sobre ML</b>	X	X	X	X							
<b>Consolidação da Base de Dados</b>		X	X								
<b>Proposição de Modelos de CNNs</b>				X	X	X	X	X			
<b>Defesa da Proposta</b>					X						
<b>Escrita do Trabalho Final</b>						X	X	X	X	X	X
<b>Treinamento das CNNs</b>					X	X	X	X	X	X	
<b>Teste das CNNs</b>					X	X	X	X	X	X	X
<b>Comparação de Métricas de Desempenho</b>						X	X	X	X	X	X
<b>Defesa do Trabalho Final</b>											X

# Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados Parciais

6. Considerações Parciais

# *Machine Learning*

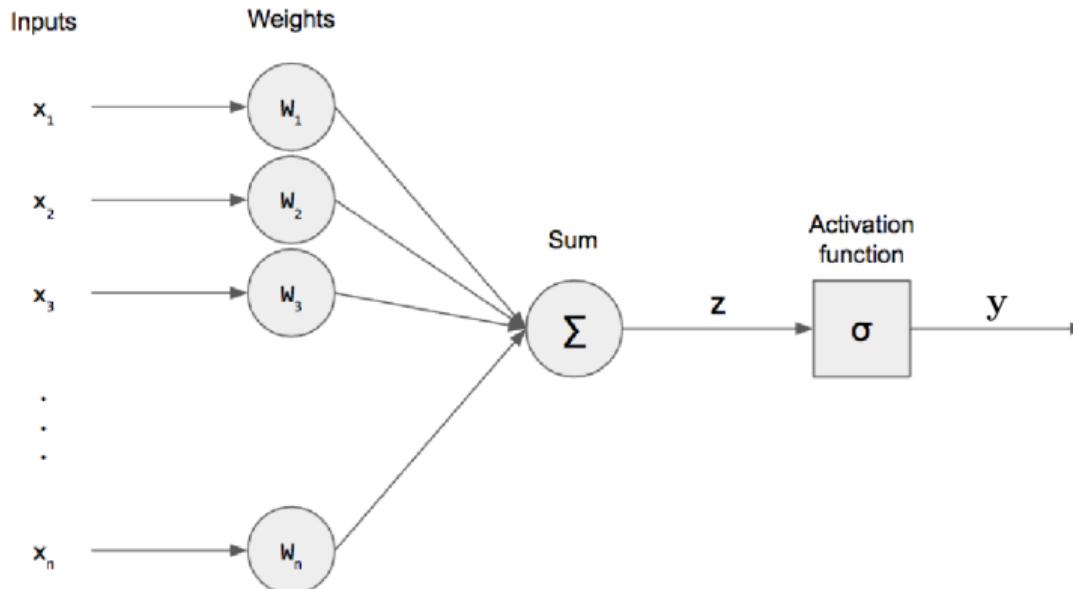
- Estudo sistemático de algoritmos e sistemas que são capazes de melhorar seu desempenho com a experiência
- Paradigmas de Aprendizado
  - Aprendizado Supervisionado
  - Aprendizado Não-Supervisionado
  - Aprendizado por Reforço
- Tarefas de Aprendizado
  - Classificação
  - Regressão

# Redes Neurais Artificiais

- Cérebro humano
- Neurônios: unidades de processamento simples
- Capacidade de capturar tendências
- Generalização

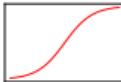
# Redes Neurais Artificiais

McCulloch e Pitts, 1943



**Figura 1:** Representação de um neurônio artificial

# Redes Neurais Artificiais

Nome	Gráfico	Equação	Intervalo
Identidade ou Linear		$\sigma(z) = z$	$(-\infty, +\infty)$
Tangente Hiperbólica		$\sigma(z) = \tanh(z) = \frac{(e^z - e^{-z})}{(e^z + e^{-z})}$	$(-1, 1)$
Sigmoide ou Logística		$\sigma(z) = \frac{1}{1 + e^{-x}}$	$(0, 1)$
Unidade Linear Retificada		$\sigma(z) = \max(0, z)$	$[0, \infty)$
Softmax		$g(z_j) = \frac{e^{z_j}}{\sum_{j=1}^K e^{z_k}} \quad j = 1, \dots, K$	$(-\infty, \infty)$

**Tabela 2:** Exemplos de funções de ativação

# Redes Neurais Artificiais

- Perceptron de Rosenblatt (1958)
  - Algoritmo de aprendizado
  - Endereçava apenas problemas linearmente separáveis

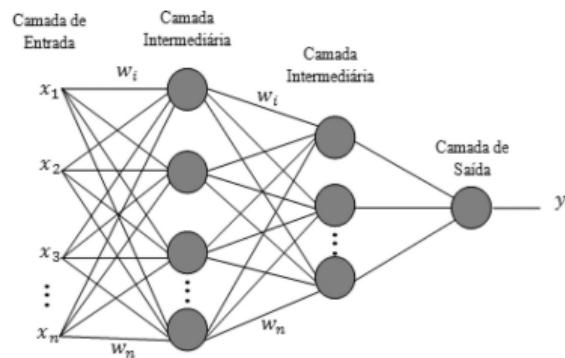
# Redes Neurais Artificiais

- Perceptron de Rosenblatt (1958)
  - Algoritmo de aprendizado
  - Endereçava apenas problemas linearmente separáveis
- Redes Neurais Artificiais
  - Organização de múltiplos neurônios artificiais sob a forma de uma rede
  - Resolução de problemas não-linearmente separáveis

# Redes Neurais Artificiais

## ► Redes Neurais *Multilayer Feedforward Perceptron*

- Camada de entrada, camadas ocultas e camada de saída
- *Feedforward* e completamente conectada
- Algoritmo *Backpropagation*
  - ▶ Fase *forward*
  - ▶ Fase *backwards*



**Figura 2:** Rede Neural MLP com duas camadas ocultas.

# Deep Learning

- Representar e reconhecer características sucessivamente complexas
- Adição de níveis ou camadas de operações não-lineares
- Resolver problemas complexos com um desempenho cada vez maior
  - Aumento recente da quantidade de dados disponíveis sobre temas complexos
  - Aumento da disponibilidade de recursos computacionais para executar modelos mais robustos
- *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*
- 14 milhões de imagens de 21 mil categorias organizadas hierarquicamente

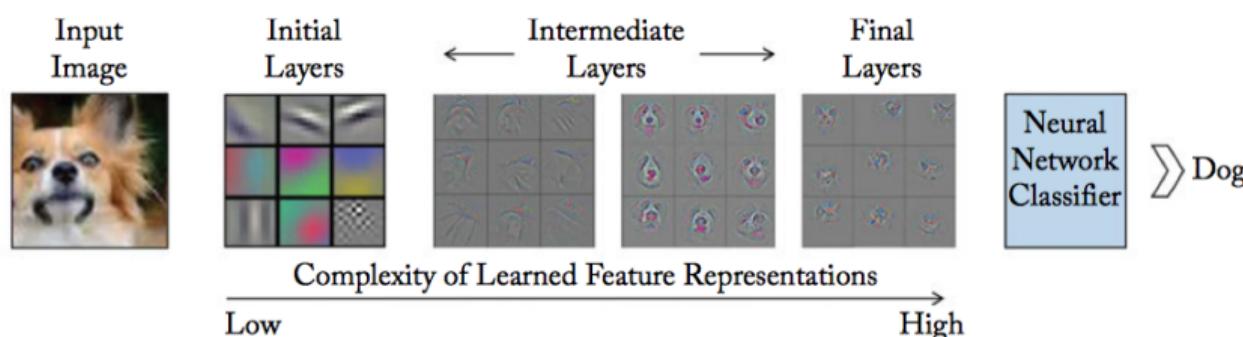
# Redes Neurais Convolucionais

- Topologia bem definida e estrutura em grid
- Destaca-se no reconhecimento de padrões em dados de alta dimensionalidade
- Diferentes tipos de camadas:
  - Camada convolucional
  - Camada de ativação
  - Camada de pooling

# Redes Neurais Convolucionais

## ▪ Convolução

$$S(i, j) = I(i, j) * K(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (1)$$



**Figura 3:** 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

# Transfer Learning

## Transfer Learning

Representações de imagens aprendidas por CNNs a partir de conjuntos de dados com grande número de exemplos podem ser transferidas eficientemente para outras tarefas de reconhecimento visual que tenham uma quantidade limitada de dados de treinamento

# Transfer Learning

- CNNs com alto desempenho
- Transferir os parâmetros de peso  $w$  e de bias  $b$
- Remove-se as últimas camadas

# Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados Parciais

6. Considerações Parciais

# Trabalhos Relacionados

## ‣ DEX (Rothe,2015)

- Problema de classificação: idade entre 0 e 100
- Dataset: *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
- Dataset: *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 Parciais

6. Considerações Parciais

# Tarefa de Aprendizado

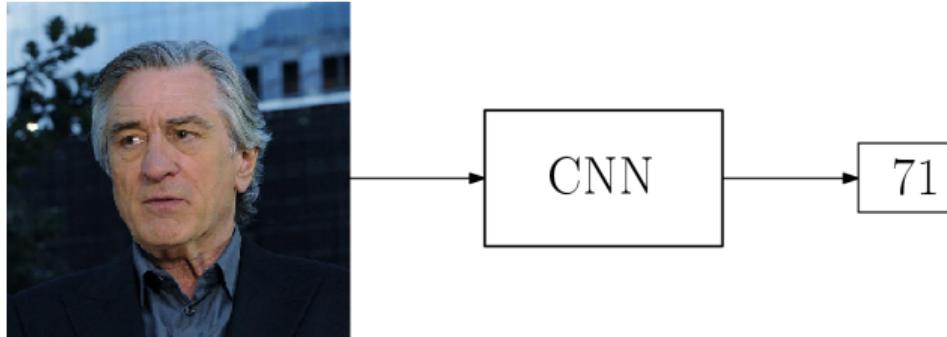
- Tarefa de Regressão

- Entrada:

- Imagem em cores RGB de dimensões  $224 \times 224$  pixels contendo uma face humana centralizada

- Saída:

- Estimativa de idade, em anos, da pessoa correspondente



# Tarefa de Aprendizado

- Métrica de desempenho: *Root Mean Squared Error* (RMSE)

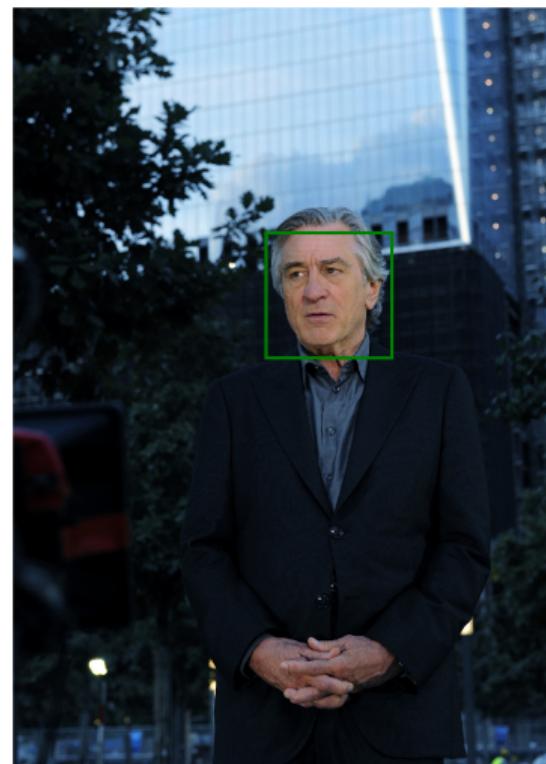
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2}. \quad (2)$$

# 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 4:** 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 5:** 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
  - $224 \times 224$  pixels
  - RGB
- Descarte de meta-dados irrelevantes para a tarefa de aprendizado

# 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 e AlexNet
- Funções de ativação tangente hiperbólica substituídas pela função *ReLU*
- Método de otimização do gradiente descendente *Adam*
- Camadas de saída com apenas um neurônio
- *batch size* igual a 64
- Número de épocas obtida de maneira experimental

# Modelos de CNN Considerados

## ➤ Primeiro ciclo de treinamento e teste

- LeNet e AlexNet
- Neurônio da camada de saída com função de ativação *ReLU*
- Taxa de aprendizado obtida de maneira experimental

## ➤ Segundo ciclo de treinamento e teste

- LeNet e AlexNet
- Neurônio da camada de saída com função de ativação *Leaky ReLU*
- Taxa de aprendizado inicial igual a  $10^{-3}$
- Decaimento de  $10^{-10}$

# Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados Parciais

6. Considerações Parciais

# Resultados Parciais

- Estimação de idade a partir de uma imagem de face
- LeNet e AlexNet

**Tabela 3:** Primeiro ciclo: Uso da função de ativação *ReLU*

Modelo	Épocas	RMSE
LeNet	95	41.08
AlexNet	55	41.96

# Resultados Parciais

- Tendência de ambas redes em preverem valores baixos
- *ReLU dying problem*
- Possível *underfitting*
- Alternativas para endereçar o problema
  - Uso da função *Leaky ReLU* na camada de saída

# Resultados Parciais

- Tendência de ambas redes em preverem valores baixos
- *ReLU dying problem*
- Possível *underfitting*
- Alternativas para endereçar o problema
  - Uso da função *Leaky ReLU* na camada de saída

**Tabela 4:** Segundo ciclo: Uso da função de ativação *Leaky ReLU*

Modelo	Épocas	RMSE
LeNet	12	41.55
AlexNet	6	14.38

# Outline

1. Apresentação

2. Fundamentação Teórica

3. Trabalhos Relacionados

4. Solução Proposta

5. Resultados Parciais

6. Considerações Parciais

# Considerações Parciais

## ➤ Resultados Preliminares

- Elaborar estratégias inteligentes para estimação de idade de telespectadores de *Smart TVs* a partir de suas respectivas fotografias faciais
- Modelos de CNNs baseados na LeNet e AlexNet
- Dois perfis de hiperparâmetros considerados
- Melhora performance verificada na AlexNet
- Em suma:
  - LeNet exibiu valores positivos e negativos próximos de zero
  - AlexNet forneceu previsões que giravam em torno da média dos dados

# Considerações Parciais

## Próximos Passos

- ▶ Pesquisar e adotar estratégias que possam minimizar os problemas identificados
  - ▶ Considerar a tarefa como um problema de classificação?
- ▶ Proposição, treinamento e teste de outras redes inspiradas em outros modelos canônicos

