

Universidade do Minho

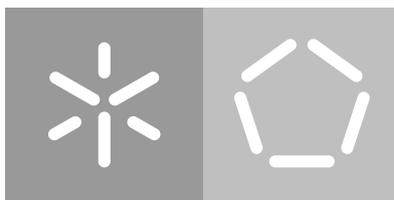
Escola de Engenharia

Departamento de Informática

Gonçalo Manuel Barbosa Teles Pinheiro

**Técnicas de Deep Learning
para a determinação da idade óssea**

December 2019



Universidade do Minho

Escola de Engenharia

Departamento de Informática

Gonçalo Manuel Barbosa Teles Pinheiro

**Técnicas de Deep Learning
para a determinação da idade óssea**

Dissertação de Mestrado

Mestrado Integrado em Engenharia Informática

Dissertação efectuada com a orientação do

Professor Doutor Miguel Angel Guevara Lopez

Professor Doutor Luís Gonzaga Mendes Magalhães

December 2019

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



<https://creativecommons.org/licenses/by/4.0/>

AGRADECIMENTOS

Em primeiro lugar agradeço aos meus orientadores o Professor Doutor Miguel Guevara e o Professor Doutor Luís Magalhães pela oportunidade de desenvolvimento deste trabalho de investigação e a sua disponibilidade e colaboração durante a sua realização.

Agradeço ao CCG - Centro de Computação Gráfica, especificamente ao CVIG, pois sem a sua estrutura a realização deste trabalho não seria possível. Desde os meus colegas, que sempre se mostraram disponíveis para esclarecer e debater os problemas encontrados, à disponibilização do *hardware* necessário para a realização desta dissertação.

Agradeço a todos os que cruzaram comigo durante o meu percurso académico e de alguma forma me influenciaram para que, com sucesso, concluísse esta etapa da minha vida.

Por último mas não menos importante agradeço à minha família, especialmente aos meus pais, por todo o apoio.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho acadêmico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

RESUMO

A avaliação da idade óssea (a maturação esquelética) é uma prática clínica comum para investigar doenças endocrinológicas, genéticas e de crescimento em crianças. Geralmente é realizada por exame radiológico da mão esquerda usando o método Greulich e Pyle (G & P) ou o Tanner Whitehouse (TW).

No entanto, ambos procedimentos clínicos demonstraram várias limitações, desde o esforço do exame que tem que ser feito pelos radiologistas até à significativa variabilidade intra e inter-operador. Para resolver este problema, várias abordagens com recurso a sistemas de apoio ao diagnóstico médico (especialmente tomando como base o método TW) foram propostas. Nenhum deles demonstrou capacidades de generalização para diferentes raças, faixas etárias e géneros.

A avaliação de exames radiológicos requer a análise de um profissional com a máxima atenção. No caso do método de Greulich e Pyle a radiografia da mão do paciente é comparada com um atlas padrão sendo possível observar deficiências no crescimento dos pacientes. Este é um trabalho exaustivo e sujeito a erros devido ao nível de atenção que é necessário durante o diagnóstico.

Os métodos de *deep learning* têm sido aplicados a diversas tarefas de análise de imagem médica como, por exemplo, classificação de lesões e segmentação de tecidos. O principal objectivo deste trabalho é desenvolver um modelo capaz de automaticamente determinar a idade óssea. Neste trabalho foram primeiramente testadas várias arquitecturas de redes neuronais convolucionais na determinação da idade óssea que mostraram bons resultados em tarefas comuns de visão por computador. Baseado nos resultados obtidos foi desenvolvido/otimizado um novo modelo que é apresentado neste documento. Foi usado *transfer learning* e o treino de raiz nas redes neuronais seleccionadas obtendo uma taxa de erro de 7,89 meses na determinação da idade óssea em pacientes do sexo feminino e uma taxa de erro de 8,28 meses ao executar esta tarefa em homens.

ABSTRACT

Bone age assessment is a common clinical practice to detect endocrinological, genetic and growth diseases. Usually it's performed using a x-ray image of the non dominant hand applying the Greulich and Pyle or the Tanner Whitehouse (TW) method.

However both procedures showed to possess several limitations since the effort demanded to radiologists to the significant intra- and inter-operator variability. To address these problems, several automated approaches (especially relying on the TW method) have been proposed. None of them showed to be able to generalize to different races, age ranges and genders.

The evaluation of x-ray images requires the analysis of a professional with maximum levels of attention. The Greulich and Pyle method consists in comparing the image with an atlas being able to observe disabilities in the growing of patients. This is an intensive job and error-prone due to the level of attention that is needed during the diagnosis.

Deep Learning methods have been applied to different medical imaging analysis tasks like, for e.g., lesion classification and tissue segmentation. The main objective of this dissertation is to develop a model capable of automatically assess bone age. In this work, we first have tested several state-of-the-art Convolution Neural Networks models for assessing bone age that previously has shown great results in general computer vision tasks. Based on these results, we have developed/optimized a new model, which is presented here. For this purpose, we used transfer learning methods and trained the selected networks from scratch achieving a 7.89-month error rate when assessing bone age in females and 8.28-month error rate when performing this task on men.

CONTEÚDO

1	INTRODUÇÃO	1
1.1	Contexto e Motivação	1
1.2	Objectivos	2
1.3	Estrutura do documento	2
2	ESTADO DA ARTE	4
2.1	Machine Learning	4
2.1.1	Redes Neurais Artificiais	5
2.1.2	Função de Custo	6
2.2	Deep Learning	7
2.2.1	Redes Neurais convolucionais	7
2.3	Visão por Computador utilizando Deep Learning	9
2.4	Idade Óssea	10
2.4.1	Anatomia da mão	10
2.4.2	Estrutura dos Ossos Longos	11
2.4.3	Indicadores de Maturidade Esquelética	12
2.4.4	Determinação Analítica da Idade Óssea	13
2.4.5	Determinação Automática da Idade Óssea	14
2.5	Computer-aided Diagnosis	15
2.6	Computer-aided Diagnosis na determinação da Idade Óssea	16
2.7	Metodos de Deep Learning na determinação da idade óssea	18
3	MATERIAL E MÉTODO	20
3.1	Material	20
3.2	Método	22
3.2.1	Pré-Processamento	24
3.2.2	Treino do Modelo	26
3.2.3	Avaliação do Modelo	26
3.2.4	Implementação	27
3.2.5	Seleccção das Arquitecturas	27
4	RESULTADOS	30
4.1	Testes das arquitecturas	30
4.2	Análise dos resultados dos testes	32
4.3	Comparação com a bibliografia	33
4.4	Generalização na determinação da idade óssea	34

5	CONCLUSÕES E TRABALHO FUTURO	35
5.1	Conclusões	35
5.2	Trabalho Futuro	36

LISTA DE FIGURAS

Figura 1	Arquitetura de um neurónio	5
Figura 2	Ilustração da aplicação do filtro à imagem	8
Figura 3	Ilustração do passo e do resultado de uma convolução	8
Figura 4	Anatomia da mão e do pulso	11
Figura 5	Estrutura do osso longo	12
Figura 6	Arquitetura do sistema <i>BoneXpert</i>	15
Figura 7	Regiões de Interesse	16
Figura 8	Metodologia proposta por Cunha et al.	17
Figura 9	Distribuições do <i>dataset</i> de Cunha et al.	21
Figura 10	Distribuições do <i>dataset</i> de <i>Radiological Society of North America (RSNA)</i>	21
Figura 11	Distribuições de idades após divisão dos <i>datasets</i>	22
Figura 12	Arquitectura definida do modelo	23

LISTA DE TABELAS

Tabela 1	Especificação do CPU e GPUs utilizados para o treino do modelo	26
Tabela 2	Detalhes das arquiteturas possíveis de escolha	28
Tabela 3	Comparação dos valores de <i>Mean Absolute Error (MAE)</i> para as diferentes percentagens aplicadas	29
Tabela 4	Tamanhos das imagens usadas para treino	30
Tabela 5	Resultados com tamanho de imagem padrão	31
Tabela 6	Resultados com o dobro do tamanho de imagem padrão	31
Tabela 7	Resultados com o quádruplo do tamanho de imagem padrão	31
Tabela 8	Resultados com o óctuplo do tamanho de imagem padrão	32
Tabela 9	Resultados da literatura	33
Tabela 10	Resultados do treino apenas com imagens do <i>dataset</i> RSNA	34
Tabela 11	Resultados da previsão das idades ósseas em pacientes do Hospital de São João	34

ACRONYMS

A

AAMS Active Appearance Models.

API Application Programming Interface.

ASM Active Shape Model.

C

CAD Computer Aided Diagnosis.

CNN Convolutional Neural Networks.

CROI Carpal Regions of Interest.

E

EMROI Epiphysis and Metaphysis Regions of Interest.

M

M₅P M₅ Model Trees.

MAE Mean Absolute Error.

MEI Mestrado em Engenharia Informática.

MLP Multi Layer Perceptron.

MSE Mean Square Error.

R

RGB Red, Green, Blue.

RMSE Root Mean Square Error.

RNA Rede Neuronal Artificial.

RNAS Redes Neuronalis Artificiais.

ROI Regions of Interest.

RSNA Radiological Society of North America.

S

SVM Support Vector Machine.

U

UM Universidade do Minho.

INTRODUÇÃO

Neste capítulo é apresentada informação sobre o contexto, a motivação e os objectivos que foram desenvolvidos nesta dissertação de mestrado. Por fim, é apresentada a estrutura do documento.

1.1 CONTEXTO E MOTIVAÇÃO

Com a evolução da tecnologia é consequente a sua aplicação nas mais diversas áreas permitindo o desenvolvimento de métodos inovadores e actualizados. Na medicina, por exemplo, é notória a sua evolução lado a lado com a tecnologia, mais concretamente na área da imagem médica. Antes do desenvolvimento de radiografias era apenas possível inferir o que se passava dentro do corpo humano através de apalpação ou observação directa.

Com esta evolução, o aparecimento de *Computer Aided Diagnosis (CAD)* é cada vez mais frequente. Os profissionais de saúde utilizam estes sistemas como uma segunda opinião de modo a tomarem decisões fundamentadas na determinação de um diagnóstico. Segundo Doi [18] a utilização de sistemas de *Computer-aided Diagnosis* é já prática comum na detecção precoce de cancro da mama em mamografias e existem já diversos sistemas no mercado, como o BoneXpert, com o objectivo de determinar a idade óssea, a partir de radiografias de mãos, mas que não se tornaram ainda uma prática comum.

Apesar desta evolução, a determinação da idade óssea continua a ser executada utilizando métodos que pouco ou nada evoluíram e não beneficiaram de tal avanço. Vários métodos utilizando técnicas clássicas de visão por computador foram propostos mas nenhum passou a ser prática comum nos *workflows* clínicos na aferição da idade óssea.

Acompanhando a evolução da inteligência artificial e o desenvolvimento de bibliotecas que permitem uma fácil utilização desta, faz todo o sentido incorporar este tipo de métodos nos processos dos sistemas de *CAD*. Em vez de se codificar à mão todos os potenciais sinais de doença, que se torna na maior parte das vezes impossível, modelos de inteligência artificial podem ser treinados para aprenderem a interpretar os resultados de exames médicos.

As abordagens existentes para a determinação automática da idade óssea podem ser divididas em três grupos: as que se baseiam unicamente na extracção e processamento de

características geométricas dos ossos da mão utilizando técnicas de visão por computador, as que utilizam modelos de *machine learning* alimentados com características geométricas dos ossos da mão e as que alimentam modelos de *deep learning* diretamente com radiografias de crianças.

A motivação para esta dissertação surge na oportunidade de aplicar tecnologias e soluções inovadoras a um problema que existe desde o século XIX e construir uma ferramenta que possa auxiliar os mais diversos profissionais de saúde na determinação da idade óssea, tornando este problema mais objectivo.

1.2 OBJECTIVOS

Considerando as motivações apresentadas anteriormente, o objectivo desta dissertação é desenvolver um método automático para a determinação da idade óssea, com recurso a técnicas de *deep learning*, capaz de auxiliar os profissionais de saúde que efectuem esta tarefa de uma forma objectiva.

Para a realização deste método são definidas cinco tarefas principais:

1. Estudar a idade óssea e os métodos utilizados para a sua determinação;
2. Estudar e seleccionar algoritmos de *Deep Learning*;
3. Implementar os algoritmos seleccionados ;
4. Testar os algoritmos seleccionados na determinação da idade óssea;
5. Analisar e discutir os resultados obtidos;

1.3 ESTRUTURA DO DOCUMENTO

Este documento está dividido em cinco capítulos que descrevem, desde, o problema da idade óssea e a sua determinação analítica, até uma solução automática para a sua resolução utilizando *deep learning*.

- O capítulo 1 apresenta os elementos de estudo, a motivação e os principais objectivos propostos para este trabalho
- O capítulo 2 introduz e contextualiza o leitor para as temáticas discutidas neste documento. Define idade óssea e apresenta os métodos utilizados actualmente na sua determinação. São apresentadas algumas noções de *machine learning* e *deep learning* e como estas se relacionam com a visão por computador. Por ultimo, são apresentados trabalhos que determinam a idade óssea recorrendo a *machine learning* e/ou *deep learning*.

- O capítulo 3 apresenta o modelo criado para abordar a determinação da idade óssea. São apresentados e analisados os dados utilizados, assim como, as diversas redes neurais convolucionais testadas para integrar o modelo.
- No capítulo 4 são discutidos os diversos testes feitos na determinação da idade óssea e os resultados obtidos nestes. É também feito um teste para perceber a capacidade de generalização do modelo em radiografias de outras regiões geográficas.
- O capítulo 5 conclui esta dissertação e apresenta os planos para trabalho futuro.

ESTADO DA ARTE

Este capítulo tem como objectivo fornecer um contexto ao trabalho apresentado nesta dissertação. São apresentados os conceitos de *machine learning* e *deep learning* com ênfase em redes neuronais convolucionais. É também apresentado o conceito de visão por computador e como este se relaciona com *deep learning* mais concretamente com redes neuronais convolucionais. De seguida, é introduzido o conceito de idade óssea e por último os trabalhos já existentes na determinação desta utilizando *deep learning*.

2.1 MACHINE LEARNING

Machine Learning é um ramo da inteligência artificial que relaciona o problema da aprendizagem a partir de dados com a inferência de conclusões generalizadas. Os processos de aprendizagem consistem em dois passos: a estimação de dependências desconhecidas de um conjunto de dados num sistema e o uso das dependências estimadas para a previsão de novos outputs do sistema [57, 96, 40, 63].

Machine Learning pode ser dividido em três vertentes de aprendizagem, supervisionada, não supervisionada e aprendizagem por reforço [53, 38, 24, 75, 32, 15, 47, 50, 89].

- Aprendizagem supervisionada ocorre quando os dados de treino estão previamente rotulados. Este tipo de aprendizagem irá permitir encontrar um modelo que seja capaz de prever um rótulo, através de regressão ou classificação, para um novo exemplo apresentado.
- Aprendizagem não supervisionada os dados de treino não estão previamente rotulados. Os métodos de aprendizagem não supervisionada consistem em explorar, ou descrever, um conjunto de dados encontrando, por exemplo, grupos de objectos semelhantes ou relacionando grupos de objectos entre si.
- Aprendizagem por reforço possibilita a um agente aprender em um ambiente controlado através de tentativa erro a partir das suas acções.

2.1.1 Redes Neurais Artificiais

Um exemplo de aprendizagem supervisionada são as *Redes Neurais Artificiais (RNAs)* que têm como inspiração o funcionamento do cérebro humano simulando a interação dos neurónios entre si. Os neurónios artificiais, ou nodos (representados na figura 1), são as unidades computacionais das RNAs. Os neurónios estão extremamente ligados e armazenam nas ligações entre si o conhecimento adquirido numa fase de aprendizagem a partir de um conjunto de dados [29, 14].

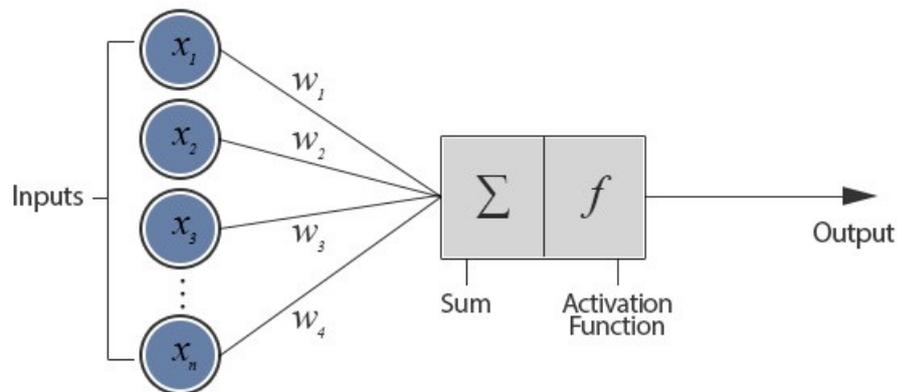


Figura 1: Arquitetura de um neurónio

Fonte: <https://medium.com/technologymadeeasy/for-dummies-the-introduction-to-neural-networks-we-all-need-c50f6012d5eb>

Acedido em 14/06/2019

Um nodo é constituído por um conjunto de conexões e uma função de activação. Cada conexão é constituída por um valor X_i e o seu peso W_i . O valor X_i representa o output de um neurónio de uma camada anterior e o peso W_i a importância dada pelo neurónio actual ao output do neurónio anterior. Os pesos serão maiores para conexões mais importantes e menores para conexões menos relevantes. As conexões são combinadas linearmente e o resultado da combinação linear das conexões é alimentada à função de activação. A função de activação tem como objectivo introduzir não linearidade no problema a ser resolvido. As funções de activação mais comuns são a *Softmax*, tangente hiperbólica ou a sigmóide.

Back-Propagation

Durante a fase de treino de uma rede neuronal artificial esta tenta ajustar-se o melhor possível de modo a reduzir o erro das previsões produzidas. A técnica mais utilizada para este efeito é o algoritmo de *back-propagation*. Tal como o nome indica, este algoritmo consiste em propagar o erro para as camadas anteriores, desde a camada de *output* até à camada de *input*, procurando o mínimo na função de custo. Este algoritmo é baseado numa técnica

chamada descida de gradiente, que usa a taxa de aprendizagem para determinar o tamanho dos passos nas suas iterações. Uma taxa de aprendizagem alta acelera a aprendizagem mas não garante que o erro seja mínimo. No entanto, com uma taxa de aprendizagem demasiado pequena o modelo demoraria um numero significativo de *epochs*(iteraões sobre o conjunto de treino) para aprender a prever novos valores. O ideal é um equilíbrio entre estes dois casos de modo a garantir que a rede neuronal é capaz de generalizar e prever em novos casos. [67]

2.1.2 Função de Custo

Para perceber se o treino de um modelo está a ser feito na direcção correcta é necessário um mecanismo para o avaliar. A função de custo, ou *loss function*, tem como objectivo avaliar a capacidade de um algoritmo em modelar os dados fornecidos. A função de custo varia consoante o tipo de problema, se se trata de classificação ou de regressão. No caso de classificação, prevê-se um output a partir de um conjunto categórico finito como, por exemplo, classificar um email como *spam* ou *ham*. Em problemas de regressão prevê se um conjunto continuo de outputs como, por exemplo, um preço de um item. No caso da regressão, quanto maior o valor pior a performance do modelo, pois traduz a distância entre os valores previstos pelo modelo e os valores reais. [100, 95]

Funções de custo de regressão

- **Mean Square Error (MSE)/Quadratic Loss/L2 Loss**

A equação 1, calcula a média do quadrado entre as distancias do valor previsto pelo modelo (f_i) e o valor real (d_i). Esta função penaliza os *outliers* devido ao facto de calcular o quadrado do desvio.

$$MSE = \frac{1}{n} \sum_{i=1}^n (d_i - f_i)^2 \quad (1)$$

- **Root Mean Square Error (RMSE)**

Presente na equação 2, esta função calcula a raiz quadrada da média do quadrado entre as distâncias do valor previsto pelo modelo (f_i) e o valor real (d_i). Tal como

mean square error, esta função penaliza os *outliers* por calcular o quadrado do desvio.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - f_i)^2} \quad (2)$$

- *MAE/ L1 Loss*

Esta função 3 calcula a média do módulo da distancia entre o valor previsto pelo modelo (f_i) e o valor real (d_i).

$$MAE = \frac{1}{n} \sum_{i=1}^n |d_i - f_i| \quad (3)$$

2.2 DEEP LEARNING

Uma *Rede Neuronal Artificial (RNA)* que possua múltiplas camadas entre a camada de entrada e a de saída denomina-se de rede neuronal profunda e caracteriza um campo de *machine learning* denominado de *deep learning* [72, 43, 76]. *Deep learning* permitiu a modelos compostos por várias camadas de processamento a aprendizagem de representações de informação. Estes métodos melhoraram consideravelmente o reconhecimento do discurso e a detecção e reconhecimento de objectos utilizando redes neuronais convolucionais [4, 61, 48, 64].

2.2.1 Redes Neuronais convolucionais

Convolução

Uma convolução consiste em percorrer uma imagem de *input* com um filtro (2a), ou *kernel*, efectuando a soma da multiplicação elemento a elemento (2b) entre o filtro e a área da imagem onde é aplicado, produzindo uma nova imagem normalmente com dimensão inferior. O tamanho da imagem depois de uma convolução é controlado pelo passo, ou *stride*. O passo indica quantos pixels se move o filtro e quanto maior o passo menor será o *output* da convolução.

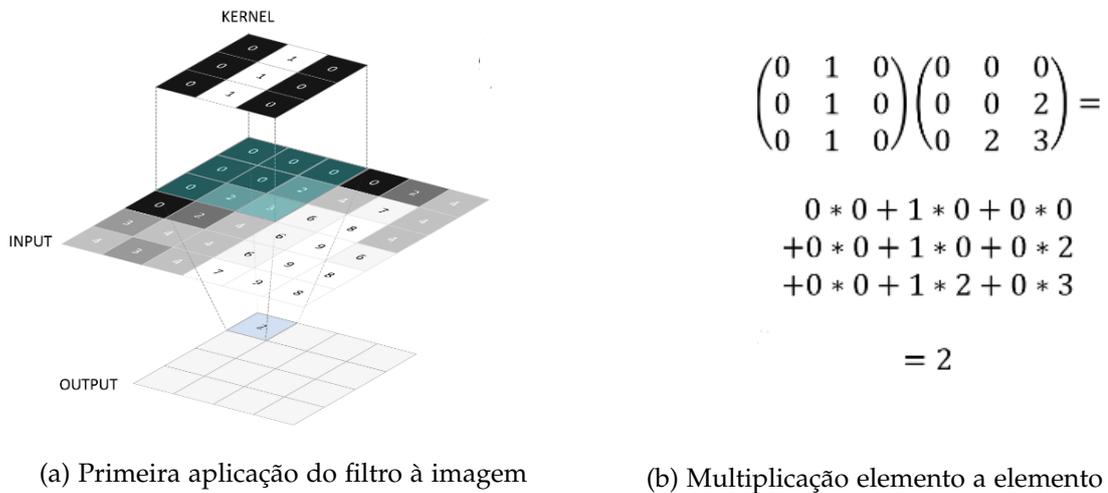


Figura 2: Ilustração da aplicação do filtro à imagem

Fonte: <https://github.com/markjay4k/Deep-Learning-Concepts/blob/master/part1-Convolution.pdf> Acedido em 14/06/2019

Na figura 3a é ilustrada a aplicação do filtro depois de um passo com tamanho um. Depois de percorrida toda a imagem dá-se por concluída a convolução sendo gerada uma nova imagem normalmente designada por *feature map*(3b).

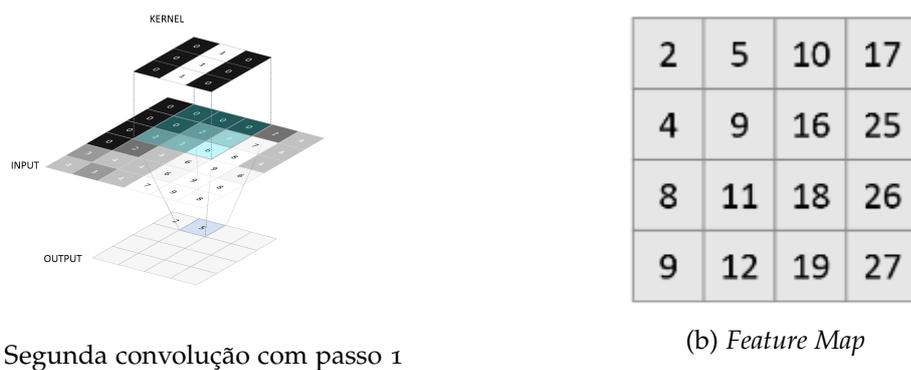


Figura 3: Ilustração do passo e do resultado de uma convolução

Fonte: <https://github.com/markjay4k/Deep-Learning-Concepts/blob/master/part1-Convolution.pdf> Acedido em 14/06/2019

Redes neuronais convolucionais (*Convolutional Neural Networks (CNN)*) processam dados que estão representados na forma de múltiplos vectores, como por exemplo uma imagem representada por três vetores de duas dimensões contendo a intensidade de cores *Red*,

Green, Blue (RGB) [43]. As principais camadas de aprendizagem que caracterizam uma CNN são a camada convolucional, a de *pooling* ou *sub-sampling* e a *fully-connected*. A primeira faz uma convolução e a camada de *pooling* gera várias amostras a partir do *feature map* gerado anteriormente. No final, através de uma camada *fully-connected*, é retirada uma conclusão seja a de pertença de uma imagem a uma classe predefinida ou a de previsão de um valor [45, 76, 11].

2.3 VISÃO POR COMPUTADOR UTILIZANDO DEEP LEARNING

Visão por computador tem como objectivo munir o computador de habilidades para compreender o mundo envolvente utilizando informação disponível em fotografias ou vídeos. Apesar de compreender as redondezas ser uma tarefa trivial para o ser humano no caso das máquinas torna-se bastante complexa. Esta complexidade deve-se ao facto de o computador armazenar as imagens como uma matriz, ou matrizes, de números. Um número destes por si só não tem qualquer significado tem por isso de ser relacionado com os seus vizinhos para ser possível retirar alguma conclusão da imagem. Existem enumeras aplicações para métodos de visão por computador como inspecção automática [83, 21], vigilância [56, 59], reconstrução de cenários [60, 73], diagnóstico médico [49, 31], entre outras.

Redes neuronais convolucionais representam o estado da arte em soluções no campo da visão por computador em tarefas como de classificação de objectos [85, 79], reconhecimento de padrões [94] e extracção de características [101]. Surgiram assim arquitecturas de CNNs que se mostraram capazes de generalizar estas tarefas como a Alexnet [41], MobileNet [35, 69], Xception [12], VGG [78], ResNet [33, 34, 97], Inception [84], DenseNet [36], NasNet [104] entre outras. A aplicação destas redes mostrou-se eficaz na compreensão de imagens médicas. Foram usadas na análise de imagens do cérebro com diferentes objectivos como a detecção de *Alzheimer*, segmentação de tecido cerebral e detecção e segmentação de lesões [25, 22]. Na imagem oftalmológica, especificamente em imagens *color fundus*, são tratados inúmeros problemas como a segmentação de estruturas anatómicas [20, 52, 103], a detecção e segmentação de anomalias na retina [8, 90] e o diagnostico de doenças [2, 31]. Em imagem torácica, seja em radiografias ou tomografias, a detecção e classificação de nódulos [51, 93, 13] e a detecção de tuberculose [70, 39] são os problemas mais trabalhados. No estudo de mamografias a tarefa mais trabalhada é a detecção de cancro que consiste em três sub-tarefas: detecção e classificação de massas [3, 6], detecção e classificação de microcalcificações [68, 58] e avaliação de risco de cancro [16]. Em imagens abdominais, como tacs ou ressonâncias magnéticas, foram utilizadas CNN na detecção e segmentação de lesões no fígado [46, 92], segmentação do pâncreas [66, 9] e a classificação de lesões na próstata [102, 5].

2.4 IDADE ÓSSEA

A idade óssea é a medida que permite definir o grau de maturação do esqueleto de uma criança ou adolescente considerando a forma, o tamanho e o grau de mineralização de um osso. Esta medida possibilita aos pediatras a detecção de doenças do foro endocrinológico, problemas de crescimento ou genéticos e também a previsão da altura adulta [30, 86, 87]. A determinação da idade óssea é efectuada a partir de uma radiografia do pulso e da mão, normalmente esquerda, aplicando um de dois métodos, o de Greulich & Pyle ou o de Tanner Whitehouse [55, 54].

Estes métodos demonstraram possuir diversas limitações, desde o esforço exigido aos radiologistas à significativa variabilidade intra e inter-operador. Os resultados destes métodos irão sempre depender do género e etnia do paciente em questão, visto que estes factores fazem variar a velocidade de desenvolvimento ósseo [17, 27, 80].

2.4.1 *Anatomia da mão*

A mão (figura 4) é constituída por ossos do carpo, metacarpo e dedos. O carpo é constituído por 8 ossos e pode ser dividido em distal e proximal. O carpo proximal está em contacto com o antebraço e é constituído pelo Escafóide, Semilunar, Piramidal e Pisiforme. O carpo distal está em contacto com o metacarpo e é constituído pelo Trapézio, Trapezóide, Capitato e Hamato. O metacarpo constitui a palma da mão e é constituído por 5 ossos. Os dedos são constituídos por falanges distais, médias e proximais com exceção do polegar que possui apenas falange proximal e distal [7].

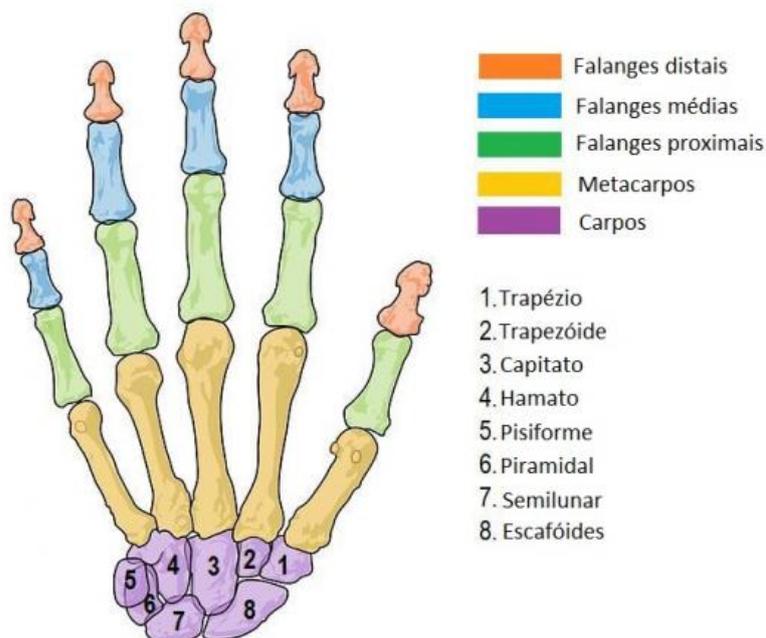


Figura 4: Anatomia da mão e do pulso

Fonte: <https://www.passeidireto.com/arquivo/36741871/ossos-da-mao> Acedido em 16/06/2019

2.4.2 Estrutura dos Ossos Longos

Ossos longos são aqueles que possuem um comprimento maior que a sua largura e são constituídos por um corpo e duas extremidades. Na mão, as falanges e os metacarpos são constituídos por ossos longos. As partes de um osso longo, observáveis na Figura 5, são:

- **Diáfise:** é o corpo de um osso longo. É constituída por tecido ósseo compacto responsável por fornecer resistência ao osso.
- **Epífise:** é a extremidade de um osso longo. Esta é responsável por unir um osso longo a uma articulação. É constituída por osso esponjoso revestida por osso compacto envolvidos em cartilagem.
- **Metáfise:** é a parte dilatada da diáfise mais próxima da epífise.

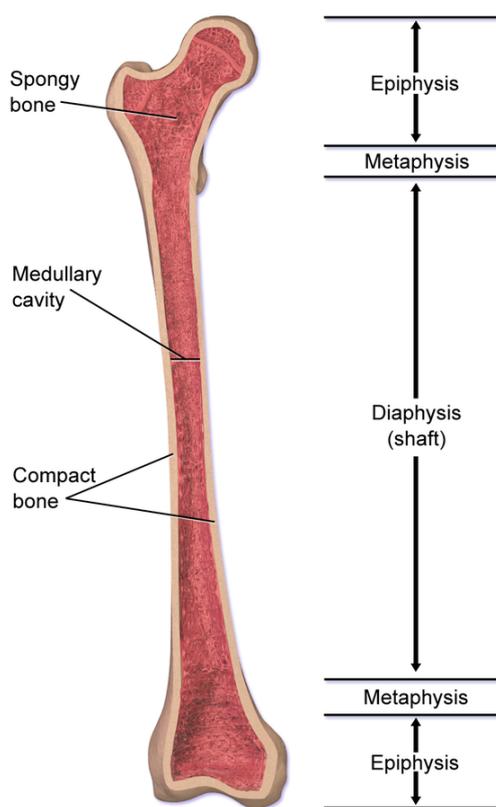


Figura 5: Estrutura do osso longo

Fonte: https://www.wikiwand.com/pt/Osso_longo Acedido em 16/06/2019

2.4.3 Indicadores de Maturidade Esquelética

Na maioria das crianças e adolescentes é possível observar, nos dois gêneros, uma ossificação constante do carpo, metacarpo e falanges. No geral, a primeira ossificação a aparecer no pulso e na mão é o capitato e, geralmente, a última ossificação é o sesamóide do polegar [23]. Para uma determinação facilitada da idade óssea, o desenvolvimento ósseo pode ser dividido em seis categorias caracterizadas por diferentes níveis de ossificação [26]:

- Infância - *Infancy*

A esta categoria pertencem crianças do sexo feminino entre o nascimento e os 10 meses e crianças do sexo masculino entre o nascimento e os 14 meses. É caracterizada pelo início da ossificação de alguns ossos do pulso e pelo aparecimento de uma pequena ossificação na epífise do rádio.

- Infância - *Toddlers*

A esta categoria pertencem crianças do sexo feminino entre os 10 meses e os 2 anos de idade e crianças do sexo masculino entre os 14 meses e os 3 anos de idade. É caracterizada principalmente pela ossificação das falanges e dos metacarpos.

- Pré-puberdade

A esta categoria pertencem crianças do sexo feminino entre os 2 e os 7 anos de idade e crianças do sexo masculino entre os 3 e os 9 anos de idade. É caracterizada pelo alargamento das epífises das falanges. Dá-se também o aparecimento de algumas ossificações nos ossos do carpo que devido à sua inconsistência não são tão confiáveis como indicadores da idade óssea.

- Início e durante puberdade

A esta categoria pertencem crianças do sexo feminino entre os 7 e os 13 anos de idade e crianças do sexo masculino entre os 9 e os 14 anos de idade. Nesta fase continua o alargamento das epífises das falanges tornando-se maiores do que as metáfises. No carpo, o pisiforme torna-se reconhecível.

- Final da puberdade

A esta categoria pertencem crianças do sexo feminino entre os 13 e os 15 anos de idade e crianças do sexo masculino entre os 14 e os 16 anos de idade. Esta fase é caracterizada pelo início da fusão das epífises às metáfises nas falanges e no metacarpo, apesar da fusão no metacarpo ser de difícil deteção em radiografias.

- Pós-puberdade

A esta categoria pertencem crianças do sexo feminino entre os 15 e os 17 anos de idade e crianças do sexo masculino entre os 16 e os 19 anos de idade. Durante esta fase a fusão das epífises às metáfises fica concluída no carpo, metacarpo e falanges.

2.4.4 *Determinação Analítica da Idade Óssea*

Nesta secção são descritos os métodos apresentados anteriormente, utilizados por radiologistas na aferição da idade óssea.

Greulich & Pyle

O método de Greulich & Pyle consiste na comparação de uma radiografia com um conjunto de radiografias que representam a normalidade para um dado grau de maturação. Contém imagens de referência de homens e mulheres. Este método foi desenvolvido através do uso, como referência, de imagens da classe média-alta de pacientes naturais da cidade de Cleveland dos Estados Unidos da América obtidas entre 1931 e 1942 [30].

Utilizando este método, o cálculo da idade óssea pode tornar-se difícil e impreciso visto que atualmente a população jovem começa a puberdade mais cedo que os sujeitos analisados durante o desenvolvimento dos atlas utilizados [71].

Tanner Whitehouse

O método de Tanner Whitehouse consiste em pontuar regiões de interesse, como a mão e o pulso, onde a soma de todas as regiões irá resultar no nível de maturação óssea geral calculando posteriormente a idade óssea [86, 87, 71]. Neste método é possível considerar três sistemas de pontuação:

- O TW 20 bones que considera, tal como o nome indica, vinte ossos sendo estes do primeiro, terceiro e quinto dedo e também os ossos do carpo,
- O RUS que considera todos os anteriores exceto os ossos do carpo
- o CARPAL que considera apenas os ossos do carpo.

O método *Tanner Whitehouse 2* foi desenvolvido utilizando radiografias recolhidas entre 1950 e 1960 de crianças da classe média-alta [86]. Posteriormente foi desenvolvido o método *Tanner Whitehouse 3* para atualizar o resultado entre a pontuação total obtida, o nível de maturação óssea e a idade óssea tendo em conta o desenvolvimento dos jovens atualmente [87].

2.4.5 *Determinação Automática da Idade Óssea*

Nesta secção é apresentado o método automático presente no mercado. A sua utilização tornou-se comum na determinação da idade óssea em hospitais e clínicas utilizando radiografias sendo que, no entanto, são também utilizados outros métodos que não recorrem a radiografias [55, 62].

BoneXpert

Este método analisa imagens de uma forma automática e independentemente da posição da mão, aceitando imagens das duas mãos. Analisa treze ossos sendo estes: o rádio, ulna e os dedos um, três e cinco. Este sistema está dividido em três camadas, observável na figura 6.

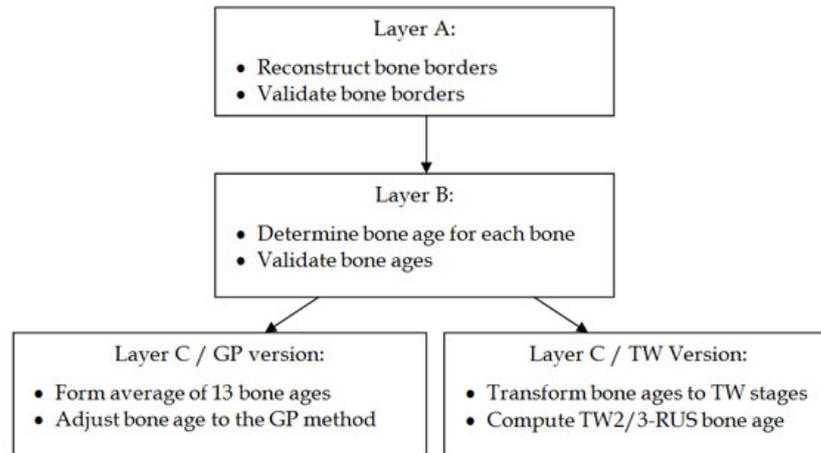


Figura 6: Arquitetura do sistema *BoneXpert*

Fonte: <https://bit.ly/2kdUfms> Acedido em 09/04/2019

A primeira camada reconstrói as bordas dos ossos utilizando *Active Appearance Models (AAMs)*. Esta reconstrução é feita moldando o modelo até coincidir com a imagem observada. A segunda camada determina a idade óssea para cada osso considerando características geométricas dos ossos em análise. Na terceira camada existem duas versões: uma aplica o método de *Greulich & Pyle* e a outra o de *Tanner Whitehouse* [88].

2.5 COMPUTER-AIDED DIAGNOSIS

Sistemas *CAD* consistem em considerar igualmente o trabalho dos profissionais de saúde e do computador. O desempenho do computador não tem de ser comparável ou melhor que o do profissional de saúde, mas complementar o seu diagnóstico [19, 18]. Na radiologia em específico, visto que o trabalho dos radiologistas se baseia numa avaliação subjetiva, o output do computador ajuda a tornar a tarefa mais objetiva [91].

Sistemas de *CAD* para radiologia são constituídos geralmente por três componentes básicos. O primeiro consiste no processamento de imagem para aprimoramento e extração de regiões de interesse. Este processamento de imagem tem como objetivo facilitar o computador, e não o humano, para detecção de possíveis lesões ou padrões anormais. O segundo componente é responsável pela quantificação das características da imagem, como tamanho, contraste e forma, realçadas no primeiro passo. No entanto, numa fase inicial do desenvolvimento de sistemas *CAD*, é útil ter definidas características que tenham sido identificadas e descritas por radiologistas devido à sua experiência e precisão na avaliação de radiografias. O terceiro passo é o processamento dos dados para a distinção de padrões normais e anormais baseado nas características obtidas no segundo componente. Este passo inclui

técnicas como métodos baseados em regras, redes neuronais artificiais e árvores de decisão. [19, 91, 62]

Técnicas de *deep learning* foram introduzidas na análise da imagem médica com resultados promissores em várias aplicações, como o prognóstico de *Alzheimer* [82], detecção de cancro [10] e a segmentação e detecção de órgãos [98].

2.6 COMPUTER-AIDED DIAGNOSIS NA DETERMINAÇÃO DA IDADE ÓSSEA

Giordano et al. [28] apresentou um sistema de avaliação da idade óssea baseado no método *Tanner Whitehouse 2*. É constituído por duas partes distintas identificadas na figura 7: a primeira processa as epífises e as metafises dos metacarpos (*Epiphysis and Metaphysis Regions of Interest (EMROI)*) enquanto a segunda processa os ossos do carpo juntamente com o rádio e ulna (*Carpal Regions of Interest (CROI)*).

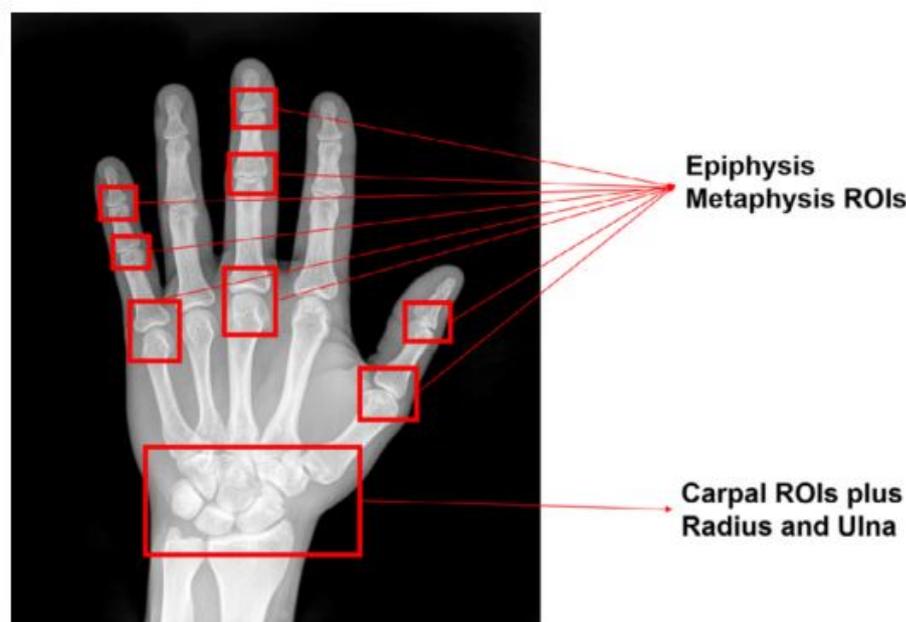


Figura 7: Regiões de Interesse

Fonte: [28]

Previamente à análise das regiões de interesse, a mão e o pulso são segmentados e são extraídas as regiões de interesse presentes na figura 7. A partir destas regiões são retiradas informações geométricas e aplicado o método tanner whitehouse. Este sistema foi testado em cento e seis radiografias, de rapazes com idades compreendidas entre os zero e os dez anos e raparigas com idades compreendidas entre os zero e os sete anos, apresentando uma taxa média de erro de 0,46 anos.

Cunha et al. [17] utilizou um *dataset* de oitocentas e oitenta e sete radiografias onde cada registo continha a idade cronológica, o género e a idade óssea estimada por radiologistas através do método de *Greulich & Pyle*. Antes de inferir a idade óssea são efetuados cinco passos como é possível observar na figura 8: Segmentação da mão, deteção dos dedos, deteção das regiões de interesse (*Regions of Interest (ROI)*), segmentação das articulações e extração de características.

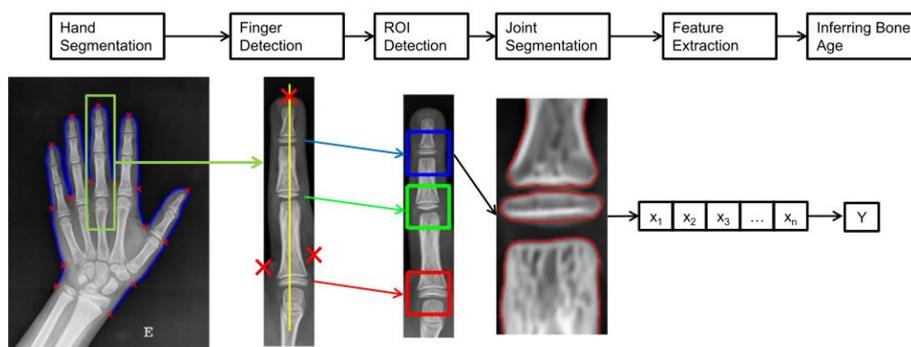


Figura 8: Metodologia proposta por Cunha et al.

Fonte: [17]

A segmentação da mão é feita utilizando um modelo denominado de *Active Shape Model (ASM)*. ASMs são modelos estatísticos que são deformados iterativamente para assentarem no objeto de outra imagem. Este modelo foi construído utilizando um conjunto de imagens com o contorno da mão realizado à mão onde foram também identificados dezasseis pontos de referência. A deteção de cada dedo é efetuada encontrando a ponta do dedo e as suas bases, direita e esquerda, e três pontos de referência assinalados no passo anterior. Depois de identificadas as referências de cada dedo é traçada uma linha no meio do contorno e respetivos pontos de referência. A extração de cada região de interesse é o passo seguinte. Uma região de interesse engloba as epífises/metáfises de cada articulação. A segmentação das articulações é realizada a cada ROI identificada anteriormente. É, de seguida, aplicado *clustering* utilizando o algoritmo *k-means* duas vezes. A primeira com três *clusters* para diferenciar o osso, o tecido e o fundo da imagem, com o objetivo de identificar e eliminar o fundo. Este processo é repetido para uma melhor diferenciação entre o osso e o tecido. Antes de proceder à segmentação do osso são aplicadas diversas técnicas de processamento de imagem para o seu realce. A segmentação é realizada utilizando o algoritmo *Chan-Vese active contour*. São então extraídas, de cada ROI, características geométricas baseadas no trabalho de Giordano et al. [28]. Estas características do osso diferem na presença ou ausência de epífise. Uma prova desta diferença é o número de regiões obtidas no passo anterior (duas ou três), com exceção dos casos de má segmentação. Na presença de apenas duas regiões, são consideradas a largura da diáfise, o mínimo e a média do

conjunto das distâncias entre a metáfise e a diáfise, assim como duas distâncias verticais: entre a posição vertical da parte mais larga da metáfise e a sua extremidade, e entre a posição vertical da parte mais larga da metáfise e o topo da extremidade da diáfise. No caso da presença de três regiões, as características são similares, não sendo utilizada a largura da diáfise, mas a largura da epífise. Foi também efetuada uma análise dos componentes principais do modelo de distribuição que descrevia o contorno da mão que resultou em doze componentes principais, também utilizados posteriormente como características. Para a fase de experimentação foram apenas considerados os casos onde três regiões estavam presentes totalizando 421 casos. Aplicaram-se cinco técnicas de regressão: regressão linear, *M5 Model Trees (M5P)*, *Multi Layer Perceptron (MLP)*, *Support Vector Machine (SVM)* e processos gaussianos. A função de custo para avaliação das regressões foi o *Mean Absolute Error (MAE)* que representa a média da diferença absoluta entre a idade prevista pela regressão e a idade atribuída pelo radiologista. Os melhores resultados foram obtidos pelo *M5P* com *bagging* (com um *bag size* de 70%) registrando um erro absoluto médio de 10,16 meses.

2.7 METODOS DE DEEP LEARNING NA DETERMINAÇÃO DA IDADE ÓSSEA

Spampinato et al. [80] propõe três abordagens diferentes utilizando *deep learning* e redes neuronais convolucionais para a detecção da idade óssea. A primeira consistiu em utilizar redes neuronais convolucionais treinadas num *dataset* diferente do utilizado para a extração de características. Este método torna-se bastante útil visto que apenas é necessário treinar a rede da regressão responsável pela determinação da idade óssea. As redes neuronais convolucionais pré-treinadas consideradas para a extração de características foram: *OverFeat* [74], *GoogLeNet* [84] e *OxfordNet* [77]. A segunda abordagem consistiu em utilizar as redes referidas anteriormente e aplicar *transfer learning*, concretamente *fine tuning*, com o *dataset* em estudo. Esta abordagem aumentou a velocidade de treino e ajudou a prevenir *overfitting*. Extraídas as características foi aplicada uma regressão linear utilizando redes neuronais artificiais. A última abordagem consistiu na criação de uma rede neuronal convolucional, apelidada de *BoNet*, treinada de raiz a partir do *dataset* das radiografias. A vantagem deste método em relação aos anteriores é que permite adaptar a arquitetura da rede às imagens a estudar possibilitando a aprendizagem de filtros específicos em vez da adaptação de alguns mais generalizados. O *dataset* utilizado continha mil trezentas e noventa e uma radiografias divididas por género e raça. Cada radiografia tinha associada duas idades atribuídas por dois radiologistas. A rede *BoNet* foi a que se mostrou mais capaz de calcular a idade óssea obteve um *Mean Absolute Error (MAE)* de 0,79 anos (9,48 meses).

Larson et al. [42] aplicaram *fine tuning* na rede *ResNet* que mostrou bons resultados no desafio ImageNet. A esta rede foram alimentadas imagens de dois hospitais pediátricos norte americanos com idades ósseas estimadas utilizando o método de *Greulich and Pyle*. O

modelo desenvolvido foi avaliado segundo o MAE e RMSE obtendo uma taxa de erro de 0.5 (6 meses) e 0.63 anos (7.56 meses), respectivamente.

Lee et al. [44] treinam a LeNet-05, uma rede neuronal convolucional, para detetar e segmentar a mão em radiografias e de seguida aplicam *fine tuning* sobre *GoogLeNet*. Sobre as imagens foram aplicadas técnicas de *data augmentation* como rotação, redimensionamento e cisalhamento adicionando, respectivamente, sessenta e uma, cento e cinquenta e vinte e uma imagens ao *dataset* original. Este método obteve um RMSE de 0.93 anos (11.16 meses) em pacientes do sexo feminino e de 0.82 anos (9.84 meses) em pacientes do sexo masculino.

Iglovikov et al. [37] abordaram a determinação da idade óssea com redes neuronais convolucionais aplicando regressão e classificação. Para o problema de regressão as idades foram normalizadas entre -1 e 1. No problema de classificação foram criadas 240 classes para representarem as idades presentes no conjunto de dados que continha 11600 imagens. A rede foi treinada com imagens da mão e pulso, apenas pulso, apenas os dedos e um conjunto de todas as anteriores. Esta abordagem conseguiu um MAE de 7.52 meses.

Zhao et al. [99] aplicaram *deep active learning* de modo a segmentar regiões de interesse, como dedos e pulso, para, seguidamente, estimar a idade óssea utilizando *paced transfer learning*. Este método obteve um MAE de 7.66 meses.

Ren et al. [65] treinam a rede *Faster-RCNN* para desenhar o rectângulo com o tamanho mínimo capaz de conter a mão na radiografia. De seguida, é aplicado *fine tuning* sobre a arquitectura *InceptionV3* para extracção de características que são, posteriormente, alimentadas a uma camada normalização seguida de uma de *pooling* e três *fully connected* encarregues do processo de regressão responsável por fornecer a idade óssea prevista. Este método obteve uma taxa de erro, calculada através do MAE, de 5.2 meses.

MATERIAL E MÉTODO

Este capítulo começa detalhando o material utilizado no desenvolvimento desta dissertação, como as imagens foram concatenadas e posteriormente divididas. É apresentado o método começando por expor uma ideia geral da estratégia adoptada e, de seguida, são especificadas todas as decisões tomadas como o pré-processamento dos dados, as condições de treino do modelo, a sua avaliação, os detalhes de implementação e a selecção e testes feitos às arquitecturas das redes neuronais convolucionais.

3.1 MATERIAL

O conjunto de dados utilizado neste trabalho é uma combinação das imagens utilizadas por Cunha et al. [17] e as imagens disponibilizadas no desafio da [RSNA](#) para a determinação da idade óssea. O *dataset* do trabalho de Cunha et al. é constituído por oitocentas e oitenta e sete radiografias onde cada registo contém a idade cronológica, o género, a vista da radiografia (posteroanterior – PA ou anteroposterior – AP) e a idade óssea estimada por radiologistas da Faculdade de Medicina da Universidade do Porto - Hospital de São João através do método de *Greulich & Pyle*. Este *dataset* contém 502 imagens do sexo feminino e 485 imagens do sexo masculino (9b). A idade cronológica varia entre os 7.7 meses (0.64 anos) e os 238.4 meses (19.87 anos) e a idade óssea varia entre os 3 meses (0.25 anos) e os 238 meses (19.83 anos) (9a).

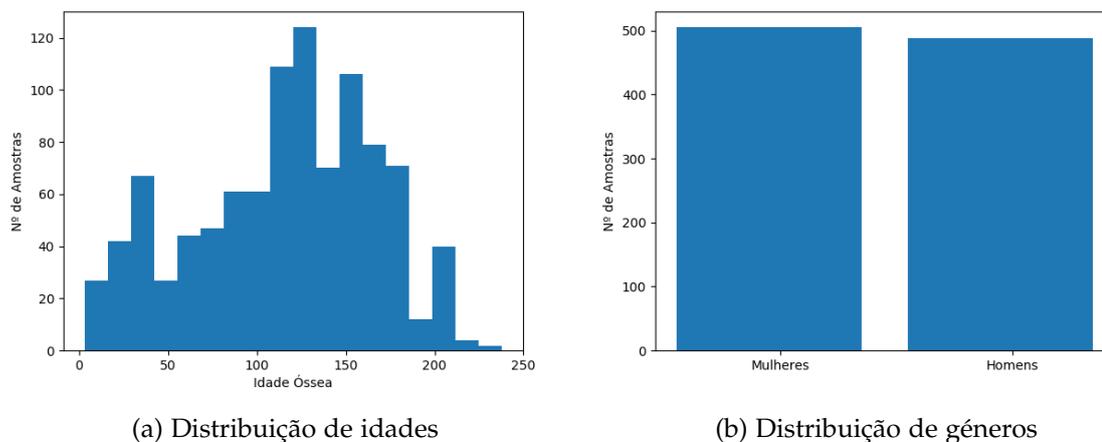


Figura 9: Distribuições do *dataset* de Cunha et al.

O *dataset* do desafio proposto pela *RSNA* é constituído por 12612 imagens doadas pelos autores de [42]. Estas imagens apenas possuem anotações sobre o género e a idade óssea. A idade óssea neste conjunto (10a) varia entre 1 mês (0.08 anos) e os 228 meses (19 anos). Existem 5778 imagens de pacientes do sexo feminino e 6833 imagens de pacientes do sexo masculino (10b).

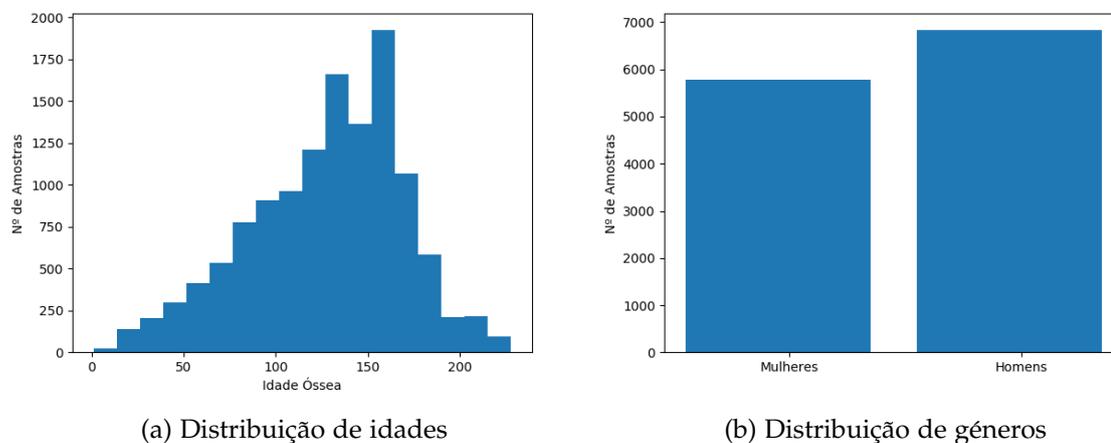
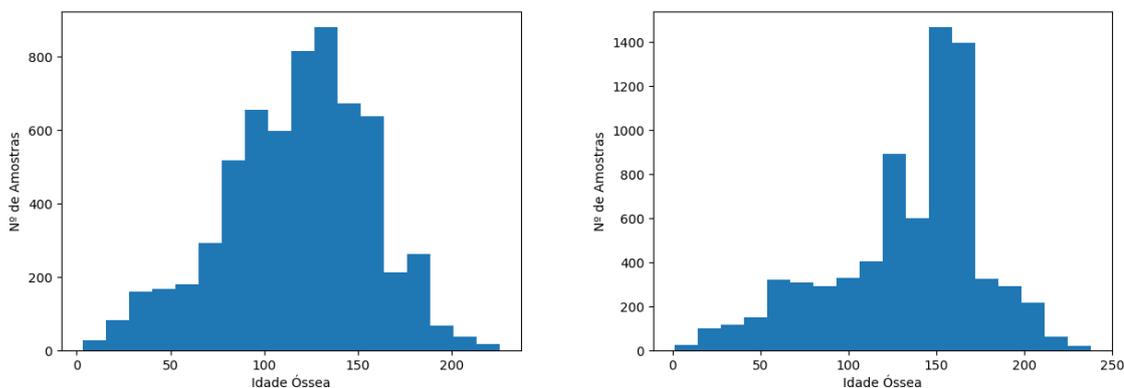


Figura 10: Distribuições do *dataset* de *RSNA*

Devido ao facto de o género ser um factor crucial na determinação da idade óssea o *dataset* foi dividido por género. Com esta divisão foram criados dois *sub-datasets* com 6283 imagens no *dataset* feminino e 7321 imagens no *dataset* masculino. Nos gráficos das figuras 11a e 11b é possível observar a distribuição das idades por género. Cada *dataset* foi dividido em 80% das imagens para treino e 20% das imagens para validação.



(a) Distribuição de idades no *dataset* feminino (b) Distribuição de idades no *dataset* masculino

Figura 11: Distribuições de idades após divisão dos *datasets*

3.2 MÉTODO

Tendo como base a bibliografia apresentada na secção 2.7, foram consideradas duas abordagens a este problema, como um problema de classificação ou como um problema de regressão. A primeira, a de classificação, consistia em dividir as idades ósseas em intervalos treinando os modelos de modo a preverem a classe pertencente de uma certa radiografia. Nesta abordagem foi detectado um desafio, que seria muito difícil de ultrapassar, que consistiu em determinar o tamanho do intervalo em que seria feita a divisão em classes das idades ósseas. A divisão em classes poderia ser feita em meses, mas o facto de as idades ósseas manterem entre si uma relação ordinal levou ao abandono desta hipótese. Foi assim decidido abordar este problema como um problema de regressão e treinar o modelo de modo a que o seu único parâmetro de saída fosse um número que representasse a idade óssea em meses correspondentes à mão contida numa radiografia.

Havia ainda uma decisão importante a tomar, que consistia em como transmitir ao modelo o género de cada imagem devido à sua importância para o resultado final. Segundo abordagens existentes isto é possível de duas formas diferentes. Uma consistia em passar ao modelo, depois da fase de extracção de características, numa das camadas *fully connected*, parâmetros indicadores do género. A segunda consistia em simplesmente dividir o *dataset* por género. Apesar da primeira abordagem dotar o modelo de uma melhor capacidade de extracção de características devido à maior quantidade de imagens observadas de uma só vez, a limitação de memória da máquina disponível levou à decisão, tal como descrito na secção 3.1, de dividir o *dataset* por género.

Considerando as decisões tomadas nos dois parágrafos anteriores, o método proposto é também baseado nas abordagens [80, 44, 1, 65]. Seguindo a abordagem dos vencedores do

desafio proposto pela [RSNA](#) para a determinação da idade óssea [1] foi decidido não adotar o passo de segmentação da mão. Por sua vez, foram aplicados outros métodos de pré-processamento, como a normalização das idades ósseas, o balanceamento da distribuição das idades ósseas e técnicas de *data augmentation*. De seguida, com base na abordagem de [80], são testadas diversas arquitecturas de redes neuronais convolucionais, que mostraram bons resultados em tarefas de classificação de objectos, para integrar o modelo final. Sobre as arquitecturas seleccionadas são efectuados testes para perceber a sua capacidade na determinação da idade óssea. Estes testes incluem a aplicação de *fine tuning* e o treino destas redes neuronais convolucionais de raiz com imagens mais detalhadas do que aquelas utilizadas na aplicação de *fine tuning*.

Sobre estas arquitecturas, baseado na implementação de [65], são adicionadas camadas de normalização, *pooling*, *dropout* e *fully connected* com o intuito, de tornar estas redes mais eficazes na determinação da idade óssea mas também transformar arquitecturas que inicialmente foram desenvolvidas para problemas de classificação de múltiplas classes num modelo de regressão. Considerando todas estas decisões a arquitectura do modelo pode ser observada na figura 12.

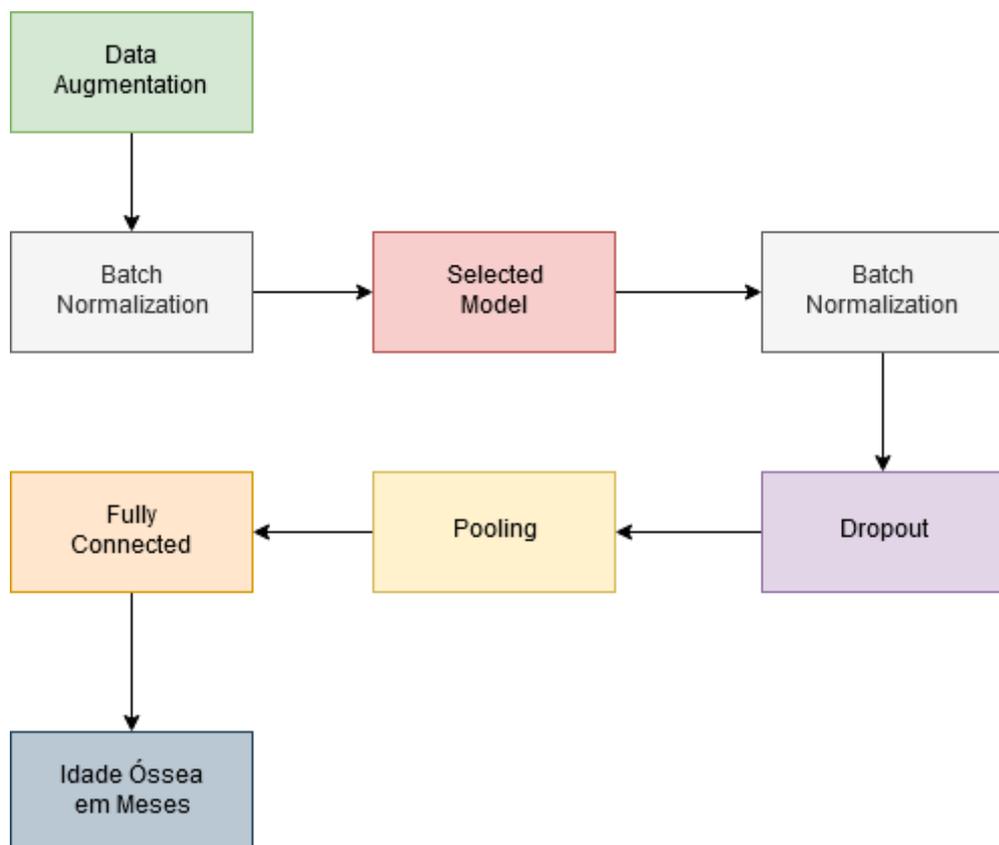


Figura 12: Arquitectura definida do modelo

3.2.1 Pré-Processamento

Normalização das idades ósseas

De modo a acelerar o processo de treino das redes neuronais todas as idades foram normalizadas segundo *z-score* calculado através da equação 4. Onde x representa a idade óssea em meses, μ a média das idades ósseas e σ o desvio padrão da amostra das idades.

$$z = \frac{x - \mu}{\sigma} \quad (4)$$

O objectivo de normalizar utilizando o *z-score* é redimensionar o intervalo das amostras preservando a distância entre estas e recolocando a média das amostras em zero e o seu desvio padrão em um. Esta normalização permite, teoricamente, que a actualização dos pesos em todos os nodos da rede neuronal seja feita à mesma velocidade e impede que alguns pesos nunca sejam actualizados ou demorem consideravelmente mais que os restantes.

Balanceamento da distribuição das idades ósseas

Depois de dividir o dataset entre teste e treino, foram criadas dez classes equidistantes, ou seja, o intervalo da amostra das idades ósseas foi dividido em dez partes iguais. De seguida, e de modo a combater o desbalanceamento entre certos intervalos de idades, algumas imagens foram repetidas em cada classe de modo a balancear as classes criadas. Este balanceamento foi aplicado em todas as classes que possuíam menos de quatrocentas imagens, preenchendo cada classe, até às quatrocentas amostras. Apesar de algumas classes possuírem mais de quatrocentas imagens e não ficarem assim perfeitamente balanceadas, este foi o valor encontrado que permitiu correr o maior número de testes com os modelos seleccionados e ter a melhor relação número de imagens / memória disponível. Ou seja, nos testes em que este número foi aumentado para além de quatrocentas imagens por classe não foi possível correr o treino das redes neuronais devido à falta de memória disponível.

Data Augmentation

De modo a aumentar a robustez e a invariância do modelo foram utilizadas técnicas de *data augmentation*. Normalmente, o processo de *data augmentation* tem como objectivo expandir o *dataset* aplicando transformações às imagens existentes e adicionando estas novas imagens ao *dataset* original. No entanto, devido às limitações de memória existentes não seria possível processar as imagens originais e as transformadas simultaneamente. A estratégia de *data augmentation* passa, então, por, a cada época, mostrar ao modelo as mesmas imagens mas com transformações diferentes permitindo que, a cada época de treino, o modelo veja novas variações das imagens originais. Adoptando esta estratégia é possível aplicar *data augmentation* sem ter um custo acrescido de memória. No início de cada treino,

são seleccionadas um número aleatório de transformações para serem aplicadas a cada imagem. A aleatoriedade das transformações permite que imagens repetidas devido ao processo descrito em 3.2.1 sejam apresentadas de forma diferente ao modelo, tendo em conta que a probabilidade de a mesma transformação ser aplicada em imagens iguais é reduzida. As técnicas aplicadas incluem: normalização de *features* ao longo de cada imagem e do dataset, *flips* horizontais, variações de altura e largura por um fator de até 10%, rotações até um máximo de 10° no sentido dos ponteiros do relógio e contra os ponteiros do relógio, cisalhamento até um máximo de 5° em ambos os sentidos e *zoom in or out* até 15%. Os parâmetros das transformações aplicadas foram estimados com base nas variações observadas entre as imagens. O processo de *data augmentation* é especificado no algoritmo 1.

Algorithm 1 Data Augmentation algorithm

```

1: procedure DATA AUGMENTATION
2:   For each epoch:
3:      $n \leftarrow \text{Random}[0, 8]$ 
4:
5:     Select random (n) from:
6:       LocalStandardization
7:       AcrossDataStandardization
8:       FlipHorizontally
9:       FlipVertically
10:      RangeHeight
11:      RangeWidth
12:      Rotate
13:      Shear
14:      Zoom
15:
16:     Apply Selected Transformations
17:
18:     Train

```

3.2.2 Treino do Modelo

O treino do modelo foi realizado na máquina especificada na tabela 1.

	CPU	GPU
Fabricante	Intel	NVIDIA
Modelo	Core i7-7700	GeForce GTX 1080
#Dispositivos	1	2
Arquitetura	Kaby Lake	Pascal
#Núcleos	4	2560
Frequência	3600	1607
Cache L1	2 x 32KB/núcleo	96KB/SM
L2 Cache	256KB/core	4MB
L3 Cache	8MB	N/A
Memória	16GB (RAM)	8GB

Tabela 1: Especificação do CPU e GPUs utilizados para o treino do modelo

De modo a tirar partido de ambos os GPUs da máquina descrita na tabela 1, é possível criar na biblioteca *Keras* um modelo *multi-GPU* que replica o modelo em ambos os GPUs aproveitando assim o paralelismo entre estes. Para este propósito, o CPU cria o modelo de modo a este ser copiado para os dois GPUs. Usando dois GPUs aumentamos para o dobro o nosso poder de processamento reduzindo, teoricamente, o tempo de treino para metade.

3.2.3 Avaliação do Modelo

A avaliação deste modelo foi feita calculando o *mean average error* (MAE) entre a idade calculada pelo radiologista e a idade prevista pelo modelo. O MAE é apenas usado como métrica e não é aplicado na optimização do modelo. A função de custo utilizada na optimização dos pesos da rede neuronal é a *mean squared error* (MSE) devido ao facto de penalizar mais os *outliers* e ter uma descida de gradiente mais suave que a MAE, permitindo uma convergência mais eficaz para o mínimo da função. O MAE mantém o gradiente durante todo o processo de treino, o que quer dizer que o gradiente se manterá elevado mesmo quando existirem valores de custo pequenos e seja necessário dar passos pequenos na função de custo. Ao contrário do MAE, o gradiente do MSE é alto quando os valores de custo são altos e diminui quando o custo tende para zero tornando o MSE mais preciso na fase final do treino.

3.2.4 Implementação

Durante o decorrer desta dissertação todo o desenvolvimento do trabalho descrito foi implementado em *Python* devido à versatilidade e oferta de bibliotecas focadas em aplicações de *Deep Learning*. Para o pré-processamento foi utilizada a biblioteca Pandas. Esta biblioteca fornece estruturas de dados e ferramentas de análise de dados que permitiram uma leitura fácil dos ficheiros CSV e a sua manipulação. A biblioteca Numpy disponibilizou um conjunto de operações matemáticas sobre vectores agilizando, por exemplo, a normalização das idades ósseas. Os gráficos de análise dos dados foram gerados através da biblioteca Matplotlib. Para o desenvolvimento do modelo de *Deep Learning* foram consideradas as seguintes bibliotecas:

- Keras - é uma *Application Programming Interface (API)* de alto nível disponibilizada em Python capaz de correr com um *backend* baseado em qualquer uma das três bibliotecas descritas de seguida. O desenvolvimento desta biblioteca teve como objectivo a rápida experimentação de modelos de *Deep Learning*.
- Tensorflow - é uma biblioteca de *Machine Learning* desenvolvida pela Google. A integração com esta biblioteca pode ser feita em Python ou C++.
- Theano - é uma biblioteca de *Machine Learning* desenvolvida pela Universidade de Monreal. A utilização desta biblioteca é feita em Python.
- Cognitive ToolKit (CNTK) - é uma biblioteca de *Machine Learning* desenvolvida pela Microsoft. A utilização desta biblioteca é feita em Python.

Das consideradas, a escolhida foi a *API* Keras utilizando como *backend* a biblioteca Tensorflow devido à facilidade com que permite a interacção com esta última e a criação e utilização de modelos já predefinidos.

3.2.5 Selecção das Arquitecturas

Neste trabalho foram testadas várias arquitecturas de redes neuronais convolucionais disponibilizadas na biblioteca *Keras* pelo módulo *keras.applications*. Esta biblioteca disponibiliza cerca de vinte arquitecturas prontas a utilizar e teria de ser adoptada uma estratégia para seleccionar cinco destas arquitecturas. As colunas *Top-1 Accuracy* e *Top-5 Accuracy* referem-se à performance destas redes no desafio de classificação *ImageNet*. *Top-1 accuracy* é a precisão da rede quando a classe prevista com maior probabilidade é de facto a classe correcta. *Top-5 accuracy* é a precisão da rede quando a classe correcta está entre as cinco previstas com maior probabilidade. Estes dois indicadores tiveram a maior relevância na escolha das arquitecturas a integrar o modelo. Ao classificarem correctamente um objecto contido numa imagem,

entende-se que são capazes de extrair características de imagens que são representativas do problema a prever e, por isso, mais capazes de generalizar. Um aspecto também considerado, com o mesmo peso, foi o tamanho, o número de parâmetros treináveis e o número de camadas de cada rede. Quantas mais camadas e parâmetros treináveis uma rede possuísse, mais memória seria necessária para carregar estes parâmetros durante a fase de *transfer learning* devido ao facto de ser necessário armazenar mais pesos para cada rede. Um último factor considerado na selecção das redes foi testar a maior variedade de arquitecturas existentes. Os detalhes de cada uma destas arquitecturas são apresentados na tabela 2.

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	26
ResNet50	98 MB	0.749	0.921	25,636,712	-
ResNet101	171 MB	0.764	0.928	44,707,176	-
ResNet152	232 MB	0.766	0.931	60,419,944	-
ResNet50V2	98 MB	0.760	0.930	25,613,800	-
ResNet101V2	171 MB	0.772	0.938	44,675,560	-
ResNet152V2	232 MB	0.780	0.942	60,380,648	-
ResNeXt50	96 MB	0.777	0.938	25,097,128	-
ResNeXt101	170 MB	0.787	0.943	44,315,560	-
InceptionV3	92 MB	0.779	0.937	23,851,784	159
MobileNet	16 MB	0.704	0.895	4,253,864	88
MobileNetV2	14 MB	0.713	0.901	3,538,984	88
DenseNet121	33 MB	0.750	0.923	8,062,504	121
DenseNet169	57 MB	0.762	0.932	14,307,880	169
DenseNet201	80 MB	0.773	0.936	20,242,984	201
NASNetMobile	23 MB	0.744	0.919	5,326,716	-
NASNetLarge	343 MB	0.825	0.960	88,949,818	-

Tabela 2: Detalhes das arquitecturas possíveis de escolha

Considerando os factores apresentados no parágrafo anterior, foram seleccionadas as seguintes arquitecturas: MobileNet [35, 69], Xception [12], VGG19 [78], InceptionV3 [84], DenseNet201 [36]. A rede MobileNet foi seleccionada devido ao seu tamanho reduzido na tentativa de perceber se uma arquitectura "simples" seria capaz de compreender e extrair características de radiografias. As restantes arquitecturas foram seleccionadas devido aos resultados já demonstrados na literatura nas mais diversas tarefas de classificação para além dos bons resultados obtidos no desafio ImageNet.

Adição de camadas

Sobre estes modelos, baseado na abordagem de [65], foram adicionadas duas camadas de *batch normalization*, uma antes para normalizar os parâmetros de entrada no modelo seleccionado e uma posteriormente para normalizar os parâmetros de saída para as próximas camadas. De seguida, é adicionada uma camada de *Global Max Pooling* com o intuito de reduzir o output do modelo e criar *feature maps*. É também adicionada uma camada de *dropout*, foram testados vários valores, demonstrados na secção seguinte, para a percentagem de *dropout* a considerar. Por ultimo, é adicionada uma camada *dense / fully connected* com o objectivo de fornecer o output final, ou seja, uma previsão em meses da idade óssea.

Seleccção de parâmetros para camada de dropout

Além de processos de *data augmentation*, camadas de *dropout* são um método eficaz e computacionalmente barato de combater *overfitting* [81]. O *dropout* consiste em remover temporariamente certos nodos da rede, juntamente com todas as suas conexões de entrada e saída, fazendo com que cada actualização numa camada durante o treino seja realizada com uma perspectiva diferente da arquitectura configurada. A percentagem de neurónios a ignorar foi estudada com base no obtido ao inferir a idade óssea com a rede *MobileNet* aplicando *transfer learning*. Foram testadas percentagens de remoção desde os 0 % até aos 50 % em intervalos de 10 %. Os melhores resultados, como é possível observar na tabela (3), foram obtidos com um *dropout* de 30% dos nodos da rede.

Dropout	MAE
0 %	13,283
10 %	11,932
20 %	10,981
30 %	10,743
40 %	12,065
50 %	13,676

Tabela 3: Comparação dos valores de MAE para as diferentes percentagens aplicadas

RESULTADOS

Neste capítulo são discutidos os resultados das diversas experiências realizadas e é feita uma comparação de resultados com os trabalhos académicos que representam o estado da arte na determinação da idade óssea. É efectuado um teste na tentativa de perceber se o modelo é capaz de generalizar a determinação da idade óssea entre pacientes de origens geográficas diferentes.

4.1 TESTES DAS ARQUITECTURAS

Cada arquitectura foi treinada sobre os pesos utilizados no desafio ImageNet com o tamanho de imagem padrão, apresentado na tabela 4, aplicando *transfer learning*. Os modelos foram também treinados de raiz com o dobro, o quádruplo e o óctuplo do tamanho padrão. Na tabela 4 estão apresentados os tamanhos das imagens de treino de cada rede. O objectivo de treinar com diferentes tamanhos de imagem foi perceber se o aumento de detalhe nas imagens estaria directamente relacionado com um melhor desempenho do modelo. Em alguns casos, o poder de computação não foi suficiente para realizar o treino devido à combinação do tamanho das imagens e do tamanho do modelo.

	Padrão	Dobro	Quádruplo	Óctuplo
Mobilenet	224x224	317x317	448x448	634x634
Xception	299x299	386x386	545x545	770x770
InceptionV3	299x299	386x386	545x545	770x770
VGG19	224x224	317x317	448x448	634x634
DenseNet201	224x224	317x317	448x448	634x634

Tabela 4: Tamanhos das imagens usadas para treino

De modo a testar o método proposto são usados os 20% de teste dos *datasets* masculino e feminino seleccionados aleatoriamente. Os resultados apresentados referem-se, em meses, ao *mean absolute error* (MAE) entre o valor inferido pelo radiologista e o valor previsto

pelo modelo. Na tabela 5 são apresentados os resultados de aplicar *transfer learning* nas arquitecturas utilizando os dois *datasets*. Normalmente, *transfer learning* implica apenas o treino das ultimas camadas de um modelo mas devido ao facto de se tratar de um problema completamente diferente do original todas as camadas são re-treinadas. Os pesos já associados a cada rede referem-se ao treino das redes na competição *ImageNet* que consiste em reconhecimento e classificação de objectos em imagens. As tabelas 6, 7 e 8 apresentam, respectivamente, os resultados dos modelos quando treinados de raiz com o dobro, o quádruplo e o óctuplo do tamanho de imagem padrão. A ausência de resultados nas tabelas sinaliza os casos em que foi impossível correr o treino devido à falta de memória para carregar imagens e o próprio modelo.

	<i>Dataset</i> Feminino	<i>Dataset</i> Masculino
Mobilenet	10.743	11.2302
Xception	7.8932	8.28
InceptionV3	8.2914	9.6882
VGG19	11.9096	14.3356
DenseNet201	9.6108	12.0708

Tabela 5: Resultados com tamanho de imagem padrão

	<i>Dataset</i> Feminino	<i>Dataset</i> Masculino
Mobilenet	10.903	13.3786
Xception	9.9122	10.9508
InceptionV3	10.6694	10.7784
VGG19	13.4516	17.1116
DenseNet201	11.5456	12.0234

Tabela 6: Resultados com o dobro do tamanho de imagem padrão

	<i>Dataset</i> Feminino	<i>Dataset</i> Masculino
Mobilenet	14.871	11.4276
Xception	16.1298	11.4496
InceptionV3	11.0214	11.895
VGG19	14.2858	15.9598
DenseNet201	-	-

Tabela 7: Resultados com o quádruplo do tamanho de imagem padrão

	<i>Dataset</i> Feminino	<i>Dataset</i> Masculino
MobileNet	11.6328	15.2742
Xception	-	-
InceptionV3	-	-
VGG19	16.3316	16.8758
DenseNet201	-	-

Tabela 8: Resultados com o óctuplo do tamanho de imagem padrão

Com base nos resultados obtidos nos testes, representados nas tabelas 5, 6, 7 e 8, a arquitectura que obteve melhores resultados foi a *Xception* quando aplicado *transfer learning*. Esta arquitectura é, então, a seleccionada para integrar o modelo de determinação da idade óssea.

4.2 ANÁLISE DOS RESULTADOS DOS TESTES

A aplicação de *transfer learning* mostrou, no geral, os melhores resultados quando comparada com o treino das redes de raiz. A rede neuronal *Xception* foi aplicada pela primeira vez na determinação da idade óssea e apresentou a menor taxa de erro, quer para o género masculino quer para o género feminino. Aplicando *transfer learning* na arquitectura *Xception* foi possível obter um MAE de 7.89 meses no dataset feminino e de 8.28 meses no dataset masculino. O facto do *transfer learning* apresentar melhores resultados que os treinos de raiz não apresenta qualquer surpresa, pois além do tempo de treino para a determinação da idade óssea estas redes possuem conhecimento prévio e tempo de treino na extracção de características, apesar de ser em um domínio diferente.

Quanto ao treino de raiz, o aumento do tamanho das imagens não se fez acompanhar, em todos os casos, de um aumento de performance por parte dos modelos. Todos os modelos quando treinados de raiz pioraram o seu desempenho em relação ao treino utilizando *transfer learning* com excepção da rede *DenseNet201* quando treinada com o *dataset* masculino. Na rede *MobileNet* o erro no dataset feminino mantém-se constante comparando *transfer learning* e o treino de raiz com o óctuplo do tamanho de imagem padrão. No entanto, a performance da rede quando treinada de raiz com imagens com o dobro e o quádruplo do tamanho de imagem padrão excede o erro do *transfer learning* com excepção do *dataset* masculino quando treinado com imagens com o dobro do tamanho padrão onde a sua performance superou o *transfer learning*. A rede *Xception*, que obteve os melhores resultados aplicando *transfer learning*, sempre que treinada com imagens mais detalhadas reduziu o seu desempenho. A rede *InceptionV3* manteve o seu desempenho muito próximo entre os treinos em que foram aplicados *transfer learning* e os treinos de raiz. A rede *VGG19* obteve em todos os casos os piores resultados de cada caso de treino mostrando-se assim

incapaz de generalizar em qualquer caso. Por ultimo, foi apenas possível treinar a rede DenseNet201 quando aplicado *transfer learning* e de raiz com o dobro do tamanho de imagem padrão devido às limitações de memória existentes. Esta rede não apresentou bons resultados em nenhum dos casos de treino.

4.3 COMPARAÇÃO COM A BIBLIOGRAFIA

Para efeitos de comparação com os restantes trabalhos da literatura foi calculada a média aritmética entre os melhores resultados obtidos nos dois *datasets* resultando num erro geral de 8.09 meses. A comparação entre trabalhos nunca poderá ser tomada em conta rigorosamente devido às radiografias utilizadas não pertencerem a pacientes das mesmas regiões e/ou raças e as estas terem sido anotadas por diferentes profissionais utilizando métodos diferentes.

	Error Rate (MAE)
Ren et al. [65]	5.2
Larson et al. [42]	6
Iglovikov et al. [37]	7.52
Zhao et al. [99]	7.66
Método Proposto	8.09
Spampinato et al. [80] Fine Tuned GoogLeNet	9.96
Spampinato et al. [80] Fine Tuned OxfordNet	9.84
Spampinato et al. [80] BoNet	9.48

Tabela 9: Resultados da literatura

Na minha opinião, o modelo não supera o desempenho de outros métodos apresentados por (1) falta de um método de segmentação da mão e (2) a impossibilidade de processar imagens com e sem transformações de uma só vez. O modelo supera o desempenho de alguns dos trabalhos apresentados na tabela 9 devido a (1) a agregação de dois *datasets* diferentes resultando em um modelo mais robusto, (2) um grande numero de transformações aplicadas às imagens durante o processo de *data augmentation* fornecendo ao modelo uma maior variedade de vistas da mão, (3) o uso da arquitectura *Xception* que nunca foi documentada em tarefas de determinação da idade óssea a partir de radiografias de mãos e (4) a adição de camadas sobre as arquitecturas testadas.

4.4 GENERALIZAÇÃO NA DETERMINAÇÃO DA IDADE ÓSSEA

Como é explicado na secção 2.4, o desenvolvimento de uma criança depende de diversos factores como a sua etnia, o género, a classe social e a sua localização geográfica. Crianças de continentes diferentes possuem graus de maturação diferentes pois estão expostas a diferentes ambientes.

Na tentativa de perceber se o modelo seria capaz de generalizar e determinar a idade óssea de pacientes de continentes diferentes foi feito um último teste. O modelo com melhor performance nos testes anteriores, o Xception, foi treinado apenas com as imagens do *dataset* do desafio *RSNA*, seguindo o processo descrito no capítulo anterior, e avaliado com o *dataset* do Hospital de São João.

O treino exclusivamente com imagens do desafio da *RSNA* obteve um *MAE* de 9.8009 meses no feminino e 11.151 meses no masculino como demonstrado na tabela 10. Na tabela 11 estão apresentados os resultados da previsão da idade óssea no *dataset* do Hospital de São João com o modelo treinado exclusivamente com imagens do desafio da *RSNA*. Esta previsão obteve um *MAE* de 19.768 meses no *dataset* feminino e de 21.3435 meses no *dataset* masculino. Este modelo mostrou-se incapaz de generalizar e de prever a idade óssea em imagens de outro continente, ou seja, mesmo utilizando técnicas de *deep learning* os pacientes de treino têm que pertencer à mesma área geográfica. Para combater esta diferença seria útil possuir, pelo menos, a raça de cada paciente e fazer esta divisão para além do género tal como o *dataset* utilizado por [80].

	Female dataset	Male dataset
Xception	9.8009	11.151

Tabela 10: Resultados do treino apenas com imagens do *dataset* *RSNA*

	Female dataset	Male dataset
Xception	19.768	21.3435

Tabela 11: Resultados da previsão das idades ósseas em pacientes do Hospital de São João

CONCLUSÕES E TRABALHO FUTURO

Neste capítulo são apresentadas as conclusões desta dissertação assim como algumas propostas para trabalho futuro.

5.1 CONCLUSÕES

Neste trabalho foi desenvolvido um método para determinar automaticamente a idade óssea em crianças e adolescentes que permitiu testar e aprimorar modelos implementados anteriormente. Ao contrário do que era pensado inicialmente, o aumento do detalhe nas imagens não foi sinónimo de melhores resultados provando-se em alguns casos o contrário. A utilização dos pesos do desafio ImageNet, ainda que com conhecimento adquirido noutra contexto, mostrou-se uma mais valia possibilitando obter os melhores resultados em todos os modelos quando aplicado *transfer learning*.

Como principal contribuição foi implementado e otimizado um modelo de redes neurais convolucionais que executou a tarefa proposta com resultados semelhantes ou melhores que métodos reportados anteriormente na literatura. Demonstrou-se que (1) o uso de técnicas de *data augmentation* consegue tornar modelos de *deep learning* mais robustos e (2) a adição de camadas específicas a modelos desenvolvidos anteriormente pode aumentar o seu desempenho. Com base no modelo proposto investigadores poderiam testar e usar outras arquitecturas de redes neurais convolucionais.

Por último, para ser possível a utilização do modelo desenvolvido em hospitais portugueses como um sistema de *computer-aided diagnosis* todo este processo teria de ser repetido utilizando apenas como referência imagens de crianças e jovens portugueses. O último teste realizado mostrou que o modelo desenvolvido quando treinado com imagens de crianças e jovens norte-americanas não é capaz de generalizar e prever correctamente a idade óssea para a população portuguesa. Idealmente, a melhor solução seria identificar diferentes regiões geográficas onde fosse possível verificar semelhanças concretas no desenvolvimento entre crianças / adolescentes.

5.2 TRABALHO FUTURO

Para trabalho futuro estão considerados os três seguintes tópicos:

- a criação de uma nova função de custo, em específico para melhorar o rácio de classificação em crianças num grau precoce de desenvolvimento (e.g., um erro de seis meses em uma criança de dois anos de idade não pode ter a mesma penalização que um erro de seis meses em um adolescente de 17 anos de idade).
- a criação de um *dataset* de pacientes portugueses que considere a sua etnia e localização geográfica.
- a criação de uma interface que facilite o trabalho de radiologistas e permita de uma maneira fácil e intuitiva a determinação da idade óssea.

BIBLIOGRAFIA

- [1] Machine learning and the future of radiology: How we won the 2017 rsna ml challenge. <https://www.16bit.ai/blog/ml-and-future-of-radiology>. Accessed: 2019-04-15.
- [2] Michael David Abràmoff, Yiyue Lou, Ali Erginay, Warren Clarida, Ryan Amelon, James C. Folk, and Meindert Niemeijer. Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning-Deep Learning Detection of Diabetic Retinopathy. *Investigative Ophthalmology Visual Science*, 57(13):5200–5206, 10 2016. ISSN 1552-5783. doi: 10.1167/iops.16-19964. URL <https://doi.org/10.1167/iops.16-19964>.
- [3] Ayelet Akselrod-Ballin, Leonid Karlinsky, Sharon Alpert, Sharbell Hasoul, Rami Ben-Ari, and Ella Barkan. A region based convolutional network for tumor detection and classification in breast mammography. In Gustavo Carneiro, Diana Mateus, Loïc Peter, Andrew Bradley, João Manuel R. S. Tavares, Vasileios Belagiannis, João Paulo Papa, Jacinto C. Nascimento, Marco Loog, Zhi Lu, Jaime S. Cardoso, and Julien Cornebise, editors, *Deep Learning and Data Labeling for Medical Applications*, pages 197–205, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46976-8.
- [4] Dario Amodei, Sundaram Ananthanarayanan, Rishita Anubhai, Jingliang Bai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Qiang Cheng, Guoliang Chen, et al. Deep speech 2: End-to-end speech recognition in english and mandarin. In *International conference on machine learning*, pages 173–182, 2016.
- [5] Shekoofeh Azizi, Farhad Imani, Sahar Ghavidel, Amir Tahmasebi, Jin Tae Kwak, Sheng Xu, Baris Turkbey, Peter Choyke, Peter Pinto, Bradford Wood, Parvin Mousavi, and Purang Abolmaesumi. Detection of prostate cancer using temporal sequences of ultrasound data: a large clinical feasibility study. *International Journal of Computer Assisted Radiology and Surgery*, 11(6):947–956, Jun 2016. ISSN 1861-6429. doi: 10.1007/s11548-016-1395-2. URL <https://doi.org/10.1007/s11548-016-1395-2>.
- [6] Maryellen L. Giger Benjamin Q. Huynh, Hui Li. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3):1 – 5 – 5, 2016. doi: 10.1117/1.JMI.3.3.034501. URL <https://doi.org/10.1117/1.JMI.3.3.034501>.

- [7] K.L. BONTRAGER and J. LAMPIGNANO. *Tratado de Posicionamento Radiografico E Anatomia*. ELSEVIER (MEDICINA). ISBN 9788535216769. URL <https://books.google.pt/books?id=Q1psznpu354C>.
- [8] P. Burlina, D. E. Freund, N. Joshi, Y. Wolfson, and N. M. Bressler. Detection of age-related macular degeneration via deep learning. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 184–188, April 2016. doi: 10.1109/ISBI.2016.7493240.
- [9] Jinzheng Cai, Le Lu, Zizhao Zhang, Fuyong Xing, Lin Yang, and Qian Yin. Pancreas segmentation in mri using graph-based decision fusion on convolutional neural networks. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 442–450, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46723-8.
- [10] J. Z. Cheng, D. Ni, Y. H. Chou, J. Qin, C. M. Tiu, Y. C. Chang, C. S. Huang, D. Shen, and C. M. Chen. Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans. *Sci Rep*, 6: 24454, 2016. ISSN 2045-2322 (Electronic) 2045-2322 (Linking). doi: 10.1038/srep24454. URL <http://www.ncbi.nlm.nih.gov/pubmed/27079888>. Cheng, Jie-Zhi Ni, Dong Chou, Yi-Hong Qin, Jing Tiu, Chui-Mei Chang, Yeun-Chung Huang, Chiun-Sheng Shen, Dinggang Chen, Chung-Ming eng Ro1 CA206100/CA/NCI NIH HHS/ Research Support, Non-U.S. Gov't England 2016/04/16 06:00 Sci Rep. 2016 Apr 15;6:24454. doi: 10.1038/srep24454.
- [11] François Chollet et al. Keras. <https://keras.io>, 2015.
- [12] François Chollet. Xception: Deep learning with depthwise separable convolutions. *CoRR*, abs/1610.02357, 2016. URL <http://arxiv.org/abs/1610.02357>.
- [13] Mark Cicero, Alexander Bilbily, Errol Colak, Tim Dowdell, Bruce Gray, Kuhan Perampaladas, and Joseph Barfett. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Investigative radiology*, 52(5):281–287, 2017.
- [14] G Coco, S Masoud-Ansari, JK Montaña Muñoz, and M Gahegan. On the use of artificial neural networks for the prediction of shoreline evolution. In *AGU Fall Meeting Abstracts*, 2018.
- [15] Alexis Conneau, Douwe Kiela, Holger Schwenk, Loic Barrault, and Antoine Bordes. Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*, 2017.

- [16] Angel Cruz-Roa, Hannah Gilmore, Ajay Basavanhally, Michael Feldman, Shridar Ganesan, Natalie NC Shih, John Tomaszewski, Fabio A González, and Anant Madabhushi. Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific reports*, 7: 46450, 2017.
- [17] P. Cunha, D. C. Moura, Guerra C. Guevara Lopez, M. A. an, D. Pinto, and I. Ramos. Impact of ensemble learning in the assessment of skeletal maturity. *J Med Syst*, 38(9):87, 2014. ISSN 1573-689X (Electronic) 0148-5598 (Linking). doi: 10.1007/s10916-014-0087-0. URL <https://www.ncbi.nlm.nih.gov/pubmed/25012476>. Cunha, Pedro Moura, Daniel C Guevara Lopez, Miguel Angel Guerra, Conceicao Pinto, Daniela Ramos, Isabel eng Research Support, Non-U.S. Gov't J Med Syst. 2014 Sep;38(9):87. doi: 10.1007/s10916-014-0087-0. Epub 2014 Jul 11.
- [18] K. Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph*, 31(4-5):198–211, 2007. ISSN 0895-6111 (Print) 0895-6111 (Linking). doi: 10.1016/j.compmedimag.2007.02.002. URL <http://www.ncbi.nlm.nih.gov/pubmed/17349778>.
- [19] Kunio Doi, Heber MacMahon, Shigehiko Katsuragawa, Robert M. Nishikawa, and Yulei Jiang. Computer-aided diagnosis in radiology: potential and pitfalls. *European Journal of Radiology*, 31(2):97–109, 1999. ISSN 0720-048X. doi: [https://doi.org/10.1016/S0720-048X\(99\)00016-9](https://doi.org/10.1016/S0720-048X(99)00016-9). URL <http://www.sciencedirect.com/science/article/pii/S0720048X99000169>.
- [20] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 698–701, April 2016. doi: 10.1109/ISBI.2016.7493362.
- [21] ES Gadelmawla. Computer vision algorithms for measurement and inspection of external screw threads. *Measurement*, 100:36–49, 2017.
- [22] Xiaohong W. Gao, Rui Hui, and Zengmin Tian. Classification of ct brain images based on deep learning networks. *Computer Methods and Programs in Biomedicine*, 138:49 – 56, 2017. ISSN 0169-2607. doi: <https://doi.org/10.1016/j.cmpb.2016.10.007>. URL <http://www.sciencedirect.com/science/article/pii/S0169260716305296>.
- [23] Stanley Marion Garn, Christabel G Rohmann, and Frederic Noah Silverman. Radiographic standards for postnatal ossification and tooth calcification. *Medical radiography and photography*, 43(2):45–66, 1967.

- [24] Christian Geiß, Patrick Aravena Pelizari, Mattia Marconcini, Wayan Sengara, Mark Edwards, Tobia Lakes, and Hannes Taubenböck. Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. *ISPRS journal of photogrammetry and remote sensing*, 104:175–188, 2015.
- [25] M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. M. van Uder, F. E. de Leeuw, E. Marchiori, B. van Ginneken, and B. Platel. Non-uniform patch sampling with deep convolutional neural networks for white matter hyperintensity segmentation. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1414–1417, April 2016. doi: 10.1109/ISBI.2016.7493532.
- [26] Vicente Gilsanz and Osman Ratib. *Hand bone age: a digital atlas of skeletal maturity*. Springer Science & Business Media, 2005.
- [27] D. Giordano, R. Leonardi, F. Maiorana, G. Scarciofalo, and C. Spampinato. Epiphysis and metaphysis extraction and classification by adaptive thresholding and dog filtering for automated skeletal bone age analysis. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6551–6556. ISBN 1094-687X. doi: 10.1109/IEMBS.2007.4353861.
- [28] Daniela Giordano, Concetto Spampinato, Giacomo Scarciofalo, and Rosalia Leonardi. An automatic system for skeletal bone age measurement by robust processing of carpal and epiphysal/metaphysal bones. *IEEE Transactions on Instrumentation and Measurement*, 59(10):2539–2553, 2010. ISSN 0018-9456 1557-9662. doi: 10.1109/tim.2010.2058210.
- [29] I Gonzalez-Fernandez, MA Iglesias-Otero, M Esteki, OA Moldes, JC Mejuto, and J Simal-Gandara. A critical review on the use of artificial neural networks in olive oil production, characterization and authentication. *Critical reviews in food science and nutrition*, pages 1–14, 2018.
- [30] W.W. Greulich and S.I. Pyle. *Radiographic Atlas of Skeletal Development of the Hand and Wrist*. Stanford University Press, 1959. ISBN 9780804703987. URL <https://books.google.pt/books?id=olezJFYxM6oC>.
- [31] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.
- [32] Moritz Hardt, Eric Price, , and Nati Srebro. Equality of opportunity in supervised learning. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and

- R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 3315–3323. Curran Associates, Inc., 2016. URL <http://papers.nips.cc/paper/6374-equality-of-opportunity-in-supervised-learning.pdf>.
- [33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. URL <http://arxiv.org/abs/1512.03385>.
- [34] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. *CoRR*, abs/1603.05027, 2016. URL <http://arxiv.org/abs/1603.05027>.
- [35] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- [36] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016. URL <http://arxiv.org/abs/1608.06993>.
- [37] Vladimir I. Iglovikov, Alexander Rakhlin, Alexandr A. Kalinin, and Alexey A. Shvets. Paediatric bone age assessment using deep convolutional neural networks. In Danail Stoyanov, Zeike Taylor, Gustavo Carneiro, Tanveer Syeda-Mahmood, Anne Martel, Lena Maier-Hein, João Manuel R.S. Tavares, Andrew Bradley, João Paulo Papa, Vasilios Belagiannis, Jacinto C. Nascimento, Zhi Lu, Sailesh Conjeti, Mehdi Moradi, Hayit Greenspan, and Anant Madabhushi, editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 300–308, Cham, 2018. Springer International Publishing. ISBN 978-3-030-00889-5.
- [38] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.
- [39] Hyo-Eun Kim and Sangheum Hwang. Scale-invariant feature learning using deconvolutional neural networks for weakly-supervised semantic segmentation. *CoRR*, abs/1602.04984, 2016. URL <http://arxiv.org/abs/1602.04984>.
- [40] Konstantina Kourou, Themis P Exarchos, Konstantinos P Exarchos, Michalis V Karamouzis, and Dimitrios I Fotiadis. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, 13:8–17, 2015.
- [41] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

- [42] David B Larson, Matthew C Chen, Matthew P Lungren, Safwan S Halabi, Nicholas V Stence, and Curtis P Langlotz. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology*, 287(1):313–322, 2017.
- [43] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–44, 2015. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature14539. URL <https://www.ncbi.nlm.nih.gov/pubmed/26017442>. LeCun, Yann Bengio, Yoshua Hinton, Geoffrey eng Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review England Nature. 2015 May 28;521(7553):436-44. doi: 10.1038/nature14539.
- [44] H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshiwas, T. K. Alkasab, G. Choy, and S. Do. Fully automated deep learning system for bone age assessment. *J Digit Imaging*, 30(4):427–441, 2017. ISSN 1618-727X (Electronic) 0897-1889 (Linking). doi: 10.1007/s10278-017-9955-8. URL <https://www.ncbi.nlm.nih.gov/pubmed/28275919>. Lee, Hyunkwang Tajmir, Shahein Lee, Jenny Zissen, Maurice Yeshiwas, Bethel Ayele Alkasab, Tarik K Choy, Garry Do, Synho eng J Digit Imaging. 2017 Aug;30(4):427-441. doi: 10.1007/s10278-017-9955-8.
- [45] S. Lee, M. Choi, H. s. Choi, M. S. Park, and S. Yoon. Fingernet: Deep learning-based robust finger joint detection from radiographs. In *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. doi: 10.1109/BioCAS.2015.7348440.
- [46] Wen Li, Fucang Jia, and Qingmao Hu. Automatic segmentation of liver tumor in ct images with deep convolutional neural networks. *Journal of Computer and Communications*, 3(11):146, 2015.
- [47] Maxwell W Libbrecht and William Stafford Noble. Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6):321, 2015.
- [48] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.
- [49] Yun Liu, Krishna Kumar Gadepalli, Mohammad Norouzi, George Dahl, Timo Kohlberger, Subhashini Venugopalan, Aleksey S Boyko, Aleksei Timofeev, Philip Q Nelson, Greg Corrado, Jason Hipp, Lily Peng, and Martin Stumpe. Detecting cancer metastases on gigapixel pathology images. Technical report, arXiv, 2017. URL <https://arxiv.org/abs/1703.02442>.
- [50] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum algorithms for supervised and unsupervised machine learning. *arXiv preprint arXiv:1307.0411*, 2013.

- [51] S. . B. Lo, S. . A. Lou, , M. T. Freedman, M. V. Chien, and S. K. Mun. Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Transactions on Medical Imaging*, 14(4):711–718, Dec 1995. ISSN 0278-0062. doi: 10.1109/42.476112.
- [52] Dwarikanath Mahapatra, Pallab K. Roy, Suman Sedai, and Rahil Garnavi. Retinal image quality classification using saliency maps and cnns. In Li Wang, Ehsan Adeli, Qian Wang, Yinghuan Shi, and Heung-Il Suk, editors, *Machine Learning in Medical Imaging*, pages 172–179, Cham, 2016. Springer International Publishing. ISBN 978-3-319-47157-0.
- [53] Stephen Marsland. *Machine learning: an algorithmic perspective*. Chapman and Hall/-CRC, 2011.
- [54] D. D. Martin, J. M. Wit, Z. Hochberg, R. R. van Rijn, O. Fricke, G. Werther, N. Cameron, T. Hertel, S. A. Wudy, G. Butler, H. H. Thodberg, G. Binder, and M. B. Ranke. The use of bone age in clinical practice - part 2. *Horm Res Paediatr*, 76(1): 10–6, 2011. ISSN 1663-2826 (Electronic) 1663-2818 (Linking). doi: 10.1159/000329374. URL <https://www.ncbi.nlm.nih.gov/pubmed/21691055>. Martin, David D Wit, Jan M Hochberg, Ze'ev van Rijn, Rick R Fricke, Oliver Werther, George Cameron, Noel Hertel, Thomas Wudy, Stefan A Butler, Gary Thodberg, Hans Henrik Binder, Gerhard Ranke, Michael B eng Research Support, Non-U.S. Gov't Review Switzerland Horm Res Paediatr. 2011;76(1):10-6. doi: 10.1159/000329374. Epub 2011 Jun 21.
- [55] Hans-J. Mentzel, Claudia Vilser, Marcus Eulenstein, Tseela Schwartz, Susanna Vogt, Joachim Böttcher, Irit Yaniv, Liat Tsoref, Eberhard Kauf, and Werner A. Kaiser. Assessment of skeletal age at the wrist in children with a new ultrasound device. *Pediatric Radiology*, 35(4):429–433, 2005. ISSN 1432-1998. doi: 10.1007/s00247-004-1385-3. URL <https://doi.org/10.1007/s00247-004-1385-3>.
- [56] Susmita A Meshram and Rani S Lande. Traffic surveillance by using image processing. In *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*, pages 1–3. IEEE, 2018.
- [57] Tom Michael Mitchell. *The discipline of machine learning*, volume 9. Carnegie Mellon University, School of Computer Science, Machine Learning . . . , 2006.
- [58] Jan-Jurre Mordang, Tim Janssen, Alessandro Bria, Thijs Kooi, Albert Gubern-Mérida, and Nico Karssemeijer. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In *International Workshop on Breast Imaging*, pages 35–42. Springer, 2016.

- [59] Naser Hossein Motlagh, Miloud Bagaa, and Tarik Taleb. Uav-based iot platform: A crowd surveillance use case. *IEEE Communications Magazine*, 55(2):128–134, 2017.
- [60] Richard A Newcombe, Dieter Fox, and Steven M Seitz. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 343–352, 2015.
- [61] Kuniaki Noda, Yuki Yamaguchi, Kazuhiro Nakadai, Hiroshi G Okuno, and Tetsuya Ogata. Audio-visual speech recognition using deep learning. *Applied Intelligence*, 42(4):722–737, 2015.
- [62] Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Lecture Notes in Computer Science. 2016. ISBN 978-3-319-46722-1 978-3-319-46723-8. doi: 10.1007/978-3-319-46723-8.
- [63] Nicolas Papernot, Patrick McDaniel, and Ian Goodfellow. Transferability in machine learning: from phenomena to black-box attacks using adversarial samples. *arXiv preprint arXiv:1605.07277*, 2016.
- [64] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [65] X. Ren, T. Li, X. Yang, S. Wang, S. Ahmad, L. Xiang, S. R. Stone, L. Li, Y. Zhan, D. Shen, and Q. Wang. Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph. *IEEE Journal of Biomedical and Health Informatics*, 23(5):2030–2038, Sep. 2019. doi: 10.1109/JBHI.2018.2876916.
- [66] Holger R. Roth, Le Lu, Amal Farag, Andrew Sohn, and Ronald M. Summers. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 451–459, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46723-8.
- [67] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.
- [68] Ravi K Samala, Heang-Ping Chan, Lubomir M Hadjiiski, Kenny Cha, and Mark A Helvie. Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis. In *Medical Imaging 2016: Computer-Aided Diagnosis*, volume 9785, page 97850Y. International Society for Optics and Photonics, 2016.

- [69] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.
- [70] Jihoon Jeong Hee-Jin Kim Sangheum Hwang, Hyo-Eun Kim. A novel approach for tuberculosis screening based on deep convolutional neural networks, 2016. URL <https://doi.org/10.1117/12.2216198>.
- [71] Mari Satoh. Bone age: assessment methods and clinical applications. *Clinical Pediatric Endocrinology*, 24(4):143–152, 2015. ISSN 0918-5739 1347-7358. doi: 10.1297/cpe.24.143. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4628949/>. 2015-0011[PII] 26568655[pmid] Clin Pediatr Endocrinol.
- [72] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.
- [73] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113, 2016.
- [74] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. *ArXiv e-prints*, December 2013.
- [75] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354, 2017.
- [76] Patrice Y Simard, David Steinkraus, and John C Platt. Best practices for convolutional neural networks applied to visual document analysis. In *ICDAR*, volume 3, pages 958–962.
- [77] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv e-prints*, September 2014.
- [78] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. URL <http://arxiv.org/abs/1409.1556>.
- [79] Richard Socher, Brody Huval, Bharath Bath, Christopher D Manning, and Andrew Y Ng. Convolutional-recursive deep learning for 3d object classification. In *Advances in neural information processing systems*, pages 656–664, 2012.
- [80] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi. Deep learning for automated skeletal bone age assessment in x-ray images. *Med Image Anal*, 36:

- 41–51, 2017. ISSN 1361-8423 (Electronic) 1361-8415 (Linking). doi: 10.1016/j.media.2016.10.010. URL <https://www.ncbi.nlm.nih.gov/pubmed/27816861>. Spampinato, C Palazzo, S Giordano, D Aldinucci, M Leonardi, R eng Netherlands Med Image Anal. 2017 Feb;36:41-51. doi: 10.1016/j.media.2016.10.010. Epub 2016 Oct 29.
- [81] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. URL <http://jmlr.org/papers/v15/srivastava14a.html>.
- [82] Heung-Il Suk, Seong-Whan Lee, and Dinggang Shen. Latent feature representation with stacked auto-encoder for ad/mci diagnosis. *Brain Structure and Function*, 220(2): 841–859, Mar 2015. ISSN 1863-2661. doi: 10.1007/s00429-013-0687-3. URL <https://doi.org/10.1007/s00429-013-0687-3>.
- [83] Da-Wen Sun. *Computer vision technology for food quality evaluation*. Academic Press, 2016.
- [84] C. Szegedy, Liu Wei, Jia Yangqing, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. ISBN 1063-6919. doi: 10.1109/CVPR.2015.7298594.
- [85] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [86] J.M. Tanner. *Assessment of skeletal maturity and prediction of adult height (TW2 method):. Tw2 Method*. Academic Press, 1983. ISBN 9780126833607. URL https://books.google.pt/books?id=_rZsAAAAMAAJ.
- [87] J.M. Tanner. *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*. Assessment of Skeletal Maturity and Prediction of Adult Height. W.B. Saunders, 2001. ISBN 9780702025112. URL <https://books.google.pt/books?id=KKdxQgAACAAJ>.
- [88] H. H. Thodberg, S. Kreiborg, A. Juul, and K. D. Pedersen. The bonexpert method for automated determination of skeletal maturity. *IEEE Transactions on Medical Imaging*, 28(1):52–66, 2009. ISSN 0278-0062. doi: 10.1109/TMI.2008.926067.
- [89] Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, and Wei-Yang Lin. Intrusion detection by machine learning: A review. *expert systems with applications*, 36(10):11994–12000, 2009.

- [90] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE Transactions on Medical Imaging*, 35(5):1273–1284, May 2016. ISSN 0278-0062. doi: 10.1109/TMI.2016.2526689.
- [91] B. Verma and J. Zakos. A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. *IEEE Transactions on Information Technology in Biomedicine*, 5(1):46–54, 2001. ISSN 1089-7771. doi: 10.1109/4233.908389.
- [92] Refael Vivanti, Ariel Ephrat, Leo Joskowicz, Naama Lev-Cohain, Onur A. Karaaslan, and Jacob Sosna. Automatic liver tumor segmentation in follow-up ct scans: Preliminary method and results. In Guorong Wu, Pierrick Coupé, Yiqiang Zhan, Brent Munsell, and Daniel Rueckert, editors, *Patch-Based Techniques in Medical Imaging*, pages 54–61, Cham, 2015. Springer International Publishing. ISBN 978-3-319-28194-0.
- [93] Changmiao Wang, Ahmed Elazab, Jianhuang Wu, and Qingmao Hu. Lung nodule classification using deep feature fusion in chest radiography. *Computerized Medical Imaging and Graphics*, 57:10 – 18, 2017. ISSN 0895-6111. doi: <https://doi.org/10.1016/j.compmedimag.2016.11.004>. URL <http://www.sciencedirect.com/science/article/pii/S0895611116301021>. Recent Developments in Machine Learning for Medical Imaging Applications.
- [94] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer, 2016.
- [95] Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.
- [96] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [97] Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *CoRR*, abs/1611.05431, 2016. URL <http://arxiv.org/abs/1611.05431>.
- [98] Wenlu Zhang, Rongjian Li, Houtao Deng, Li Wang, Weili Lin, Shuiwang Ji, and Dinggang Shen. Deep convolutional neural networks for multi-modality iso-intense infant brain image segmentation. *NeuroImage*, 108(Supplement C):214 – 224, 2015. ISSN 1053-8119. doi: <https://doi.org/10.1016/j.neuroimage.2014.12.061>. URL <http://www.sciencedirect.com/science/article/pii/S1053811914010660>.

- [99] Chen Zhao, Jungang Han, Yang Jia, Lianghui Fan, and Fan Gou. Versatile framework for medical image processing and analysis with application to automatic bone age assessment. *Journal of Electrical and Computer Engineering*, 2018, 2018.
- [100] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2016.
- [101] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.
- [102] Yulian Zhu, Li Wang, Mingxia Liu, Chunjun Qian, Ambereen Yousuf, Aytekin Oto, and Dinggang Shen. Mri-based prostate cancer detection with high-level representation and hierarchical classification. *Medical physics*, 44(3):1028–1039, 2017.
- [103] Julian Zilly, Joachim M. Buhmann, and Dwarikanath Mahapatra. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Computerized Medical Imaging and Graphics*, 55:28 – 41, 2017. ISSN 0895-6111. doi: <https://doi.org/10.1016/j.compmedimag.2016.07.012>. URL <http://www.sciencedirect.com/science/article/pii/S0895611116300775>. Special Issue on Ophthalmic Medical Image Analysis.
- [104] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition. *CoRR*, abs/1707.07012, 2017. URL <http://arxiv.org/abs/1707.07012>.

