

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Nome: Frederico Guth

CPF: 273.723.818-86

Proposta de Projeto de Pesquisa

Título: Transferência de Aprendizado em Visão Computacional

Linha de Pesquisa: Sistemas de Computação

Área de Pesquisa: Visão Computacional

Transferência de Aprendizado em Visão Computacional

Frederico Guth

1 Introdução

Recentes avanços na área de Visão Computacional tornam possíveis aplicações que vêm merecendo atenção da mídia e público: são capazes de reconhecer de pessoas, lugares e objetos com acurácia super-humana[6], diagnosticar câncer de pele tão bem quanto dermatologistas[8], segmentar semanticamente cenas em tempo real para carros autônomos, localizar tumores em imagens de ressonância magnética, ver através de paredes usando sinais de rádio, entre tantas outras. Tal avanço apresenta um contraste extremo com como a comunidade se via há apenas 10 ou 20 anos:

”Apesar de como campo de pesquisa, [Visão Computacional] apresentar problemas interessantes e desafiadores, em termos de aplicações práticas bem sucedidas é decepcionante”
—T. S. Huang, 1996 [4].

O momento crucial para tal metóricoprogresso foi o resultado de Alex Krizhevsky et al.[5] no desafio *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) de 2012 [1]. Em 8 anos de ILSVRC, o erro no reconhecimento de objetos diminuiu uma ordem de magnitude[6] e, em 2017, chegou a apenas 2,3% (menos da metade do erro humano [2]). Três desenvolvimentos simultâneos possibilitaram tal feito[3]: (a) redes convolucionais profundas, em que características visuais (*features*) são aprendidas dos dados ao invés de manualmente elaboradas; (b) barateamento do custo computacional para treinamento de algoritmos; (c) construção de bancos de imagens de larga escala com milhões de imagens e milhares de classes bem anotadas.

Neste contexto, é compreensível se iludir com o sensacionalismo e até pensar que Aprendizado Profundo (DL¹) para Visão Computacional seja uma área ”resolvida”e que agora virá um novo ”inverno”no progresso de Aprendizado de Máquina (ML²)—longe disso. Os melhores casos de sucesso foram desenvolvidos exclusivamente para uma tarefa, em um único domínio, ou seja, sobre a premissa básica que na prática, os dados com os quais o modelo será testado são do mesmo espaço de características (*feature space*) e possuem a mesma distribuição que os dados de treinamento[7]. Em outras palavras, temos algoritmos cada vez melhores que aprendem a encontrar padrões a partir de uma grande quantidade de dados rotulados, mas que ainda

¹do inglês, Deep Learning

²do inglês, Machine Learning

são deficientes na capacidade de generalizar para condições diferentes daquelas encontradas no treinamento.

Ao mesmo tempo que bancos de imagens de larga escala são um dos componentes chave para progresso, cada vez mais a necessidade de dados são um fator limitante para aplicação prática de inteligência artificial. Uma quantidade importante de comportamentos da natureza e atividades humanas obedecem a uma distribuição de cauda longa, o que torna bastante difícil a coleta representativa de dados. Há também o alto custo de rotulação: em diversas áreas de aplicação, e.g. na medicina, dados bem rotulados são extremamente difíceis de serem obtidos e pode-se levar anos para obter poucas dezenas de amostras. O mundo real é variado, os modelos encontrarão, na prática, diversos cenários para os quais não foram treinados.

Transferência de aprendizado (TL³), nos permite lidar com a insuficiência de dados para treinamento, usando conhecimento adquirido em um outro domínio ou tarefa. Essa capacidade é absolutamente necessária para o uso de inteligência artificial em larga escala que vai além das tarefas e domínios para os quais a disponibilidade de dados rotulados é abundante. Em outras palavras, apesar de todo o sucesso, os modelos atuais atacam apenas os casos fáceis em termos de disponibilidade de dados. Para lidar com um mundo cauda longa, precisamos aprender a transferir aprendizado.

1.1 Definição Formal

Transferência de Aprendizado envolve os conceitos de *domínio* e *tarefa*. Seguindo a notação de [7], um domínio \mathcal{D} é composto de um espaço de características $\mathcal{X} \subset R^d$ e uma distribuição marginal $P(X)$, onde $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. Para um problema de classificação de imagens, por exemplo, \mathcal{X} é o espaço de todas possíveis representações de imagens, com suas dimensões e canais, x_i é uma imagem e X é o *dataset* de treinamento.

Dado um domínio $\mathcal{D} = \{\mathcal{X}, P(X)\}$, uma tarefa \mathcal{T} é definida pelo espaço de rótulos \mathcal{Y} com distribuição condicional $P(Y|X)$, ou seja, $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$, onde $f(\cdot)$ é uma função objetivo que dado um $x_i \in \mathcal{D}$, prediz seu correspondente $y_i \in \mathcal{Y}$.

Dado um domínio fonte \mathcal{D}_S e uma tarefa \mathcal{T}_S , um domínio destino \mathcal{D}_T e uma tarefa \mathcal{T}_T , **transferência de aprendizado** objetiva auxiliar o aprendizado da função de predição $f_T(\cdot)$ em \mathcal{D}_T usando conhecimento de \mathcal{D}_S e \mathcal{T}_S , onde $\mathcal{D}_S \neq \mathcal{D}_T$ ou $\mathcal{T}_S \neq \mathcal{T}_T$.

2 Justificativa

O incontestado sucesso obtido nos últimos anos com aplicações de ML teve como principal motor o uso de aprendizado supervisionado em redes neurais profundas. Este paradigma é dependente da disponibilidade de dados bem rotulados que são caros e difíceis de obter. A próxima fron-

³do inglês, *Transfer Learning*

teira está em lidar com problemas para os quais há insuficiência de dados rotulados. Isso leva a especialistas da área a eleger TL como um dos campos mais promissores para o futuro:

”Transferência de Aprendizado será o próximo motor do sucesso comercial com Aprendizado de Máquinas.” —Andrew Ng, Tutorial NIPS 2016 [?]

Há um grande interesse acadêmico no assunto, a se destacar, por exemplo, que nas últimas duas edições, 2018 e 2017, da importante CVRP⁴, trabalhos sobre TL ganharam o prêmio de melhor *paper*. Entretanto, TL ainda é tratada de uma forma *ad hoc* e, apesar de alguns notáveis esforços, carece de arcabouço teórico.

Machine Learning tem como inspiração a biologia humana e estudos mais recentes demonstram correspondência entre padrões aprendidos por redes neurais convolucionais e aqueles compreendidos por animais. Entretanto, praticamente não há referência aos estudos psicológicos do aprendizado humano como inspiração para Transferência de Aprendizado e há por demais evidências que humanos têm uma enorme capacidade de generalização, de se adaptar a novos cenários. Há algo ainda incompreendido no ”*software*” humano que se reflete nessa extraordinária e eficiente capacidade de transferir aprendizado de experiências passadas para novos desafios, mas o que já sabemos sobre o aprendizado humano é bem capaz de trazer novos *insights* para ML e valem ser estudados sob este novo prisma.

3 Objetivos

O objetivo geral do presente projeto é desenvolver e investigar métodos de transferência de aprendizado para problemas de visão computacional. Espera-se criar métodos mais versáteis e eficientes, requerendo menos dados para serem treinados.

3.1 Objetivos específicos

Como objetivos específicos, pretende-se:

- Investigar e reproduzir pesquisas recentes em:
 - *self-supervised learning*;
 - *multi-task learning*;
 - *domain-adaptation*, com interesse especial em adaptação de dados sintéticos;
 - *selection bias shift*;
 - *meta-learning*;
 - *weakly-supervised learning*;

⁴Conference on Computer Vision and Pattern Recognition

- *one-shot learning* e *few-shots learning*;
- *Generative Adversarial Networks* e suas implicações para TL;
- *learning theoretic*;
- Classificar, organizar e diferenciar os diferentes métodos existentes de transferência de aprendizado;
- Estudar as principais teorias psicológicas sobre aquisição de aprendizado em humanos;
- Propor novos métodos de transferência de aprendizado;
- Publicar papers em revistas e conferências.

4 Revisão da Literatura

5 Metodologia

Não sei bem o que fazer.

6 Plano de Trabalho

As seguintes tarefas serão executadas no desenvolvimento deste projeto:

1. Atualização do Conhecimento sobre TL e áreas correlatas na Ciência da Computação:
 - (a) Revisão bibliográfica para atualização com o estado-da-arte nas áreas relacionadas ao projeto.
 - (b) Avaliação, instalação e adaptação a ferramentas de desenvolvimento que possam ser aplicadas ao projeto.
 - (c) Definição das pesquisas candidatas a serem reproduzidas.
 - (d) Relatório técnico da revisão bibliográfica de Ciência da Computação.
2. Estudo de áreas correlatas externas à Ciência da Computação:
 - (a) Desenvolvimento cognitivo em crianças.
 - (b) Cognição visual em humanos.
 - (c) Espistemologia.
 - (d) Ética
 - (e) Relatório técnico da revisão bibliográfica externa à Ciência da Computação.

3. Domain Adaptation, Synthetic datasets, GANs

- (a) Reprodução de experimentos.
- (b) Relatório Técnico.

4. Primeiro Paper

- (a) Projeto
- (b) Desenvolvimento
- (c) Avaliações Experimentais
- (d) Análise
- (e) Escrita

5. Mult-task Learning, Self-supervised Learning

- (a) Reprodução de experimentos.
- (b) Relatório Técnico.

6. One-Shot Learning, Few-Shots Learning, Weakly Supervised Learning

- (a) Reprodução de experimentos.
- (b) Relatório Técnico.

7. Segundo Paper

- (a) Projeto
- (b) Desenvolvimento
- (c) Avaliações Experimentais
- (d) Análise
- (e) Escrita

8. Learning Theoretic, Meta-Learning

- (a) Reprodução de experimentos.
- (b) Relatório Técnico.

9. Primeiro esboço da Tese

- (a) Escrita

10. Terceiro Paper

- (a) Projeto

- (b) Desenvolvimento
- (c) Avaliações Experimentais
- (d) Análise
- (e) Escrita

11. Tese

- (a) Escrita

7 Cronograma

Fazer. A ideia é caber em 3 anos e meio, sendo que de 3 a 10, são 8 tarefas que eu estou imaginando levar 3 a 4 meses cada. De forma que se forem 4 meses, serão 32 meses, 2 anos e meio. As tarefas 1 e 2 serão desenvolvidas ao longo dos anos. E a tarefa 11 deve durar um 6 meses.

Bibliografia

- [1] Ian J. Goodfellow, Yoshua Bengio, and Aaron C. Courville. *Deep Learning*. Adaptive computation and machine learning. MIT Press, 2016.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *CoRR*, abs/1502.01852, 2015.
- [3] Grant Van Horn and Pietro Perona. The devil is in the tails: Fine-grained classification in the wild. *CoRR*, abs/1709.01450, 2017.
- [4] T Huang. *Computer vision: Evolution and promise*, 1996.
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [6] J. Deng L. Fei-Fei. Where have we been? where are we going?, 2017. [Online; accessada 28 de Junho de 2018].
- [7] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Trans. on Knowl. and Data Eng.*, 22(10):1345–1359, October 2010.

- [8] Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 5987–5995, 2017.