Transfer Learning

Tutorial at SIBGRAPI 2019

Ricardo Ribani
Mackenzie
HP

Mauricio Marengoni
Mackenzie

Mackenzie

Sponsorship:

Mack Pesquisa

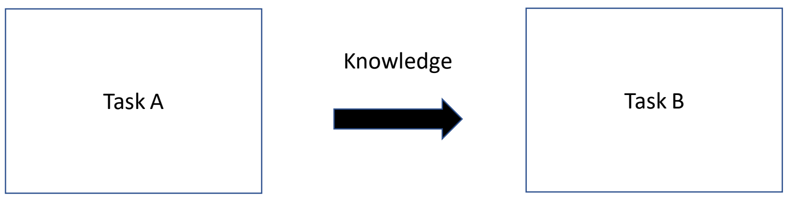
Tutorial Program and Schedule

- **Part I**
  - Introduction
  - Overview and Definition
  - Transfer Learning Strategies
  - Approaches
  - Deep Learning
  - Types of Transfer Learning
  - Conclusion and Future
- *Break (coffee)*
- **Part II Testing**
  - Practical Use Cases

Introduction

What is transfer learning ?

In general the idea of transfer learning is to use knowledge acquired for one task to solve related ones.

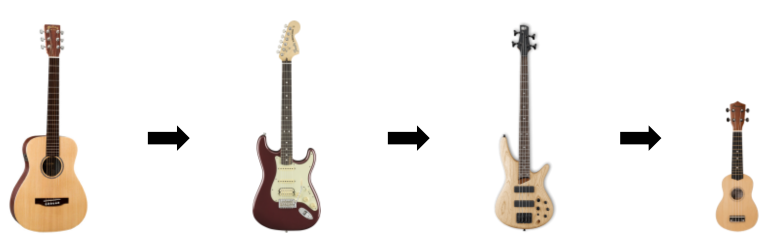


Introduction

"Humans have an inherent ability to transfer knowledge across tasks. The more related the tasks, the easier it is for us to transfer, or cross-utilize our knowledge."
D. Sarkar (2018)

Example:

Play Guitar → Learn to play Electric Guitar → Learn to play Bass → Learn to play Ukelele



Introduction

Another Example:

Know how to Skateboard → Learn to Surf

➡

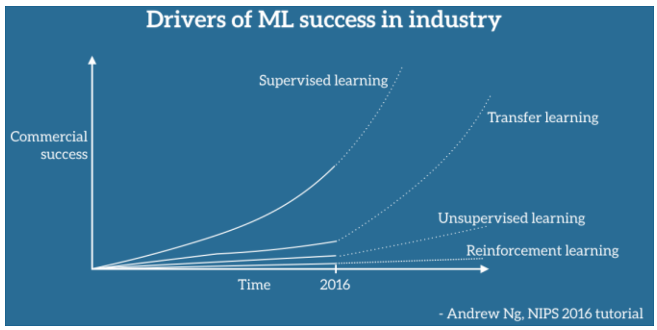
Introduction

The idea that people can apply knowledge already learned from one task to another, faster and with better success, led to the study of transfer learning.

This motivation was discussed in the NIPS-1995 workshop “Learning to Learn” in order to discuss progress in knowledge consolidation and transfer.

Introduction

Recently reinforced by Andrew Ng in NIPS-2016 "Nuts and bolts of building AI applications using Deep Learning", who mentioned that transfer learning will be the next driver of machine learning commercial success.



Introduction

The idea of transfer learning has been discussed in the literature since 1990's with different names:

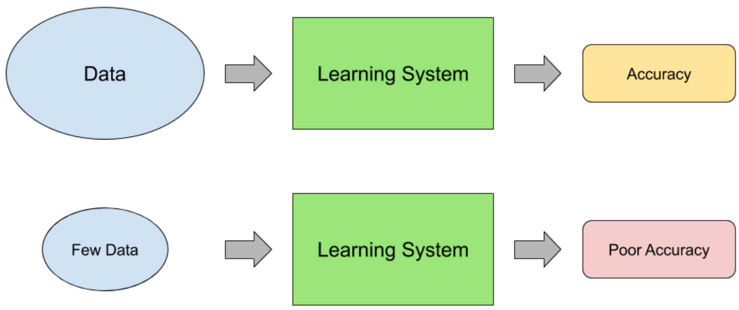
- Learning to learn
- Life-long learning
- Knowledge transfer
- Inductive transfer
- Multi-task learning
- Knowledge consolidation
- Context-sensitive learning
- Knowledge-based inductive bias
- Meta-learning
- Cumulative learning

All of these methods aim to extract knowledge from one or more tasks and apply to a different new task.

Introduction

Why to use transfer learning ?

Supervised learning methods need labeled data!



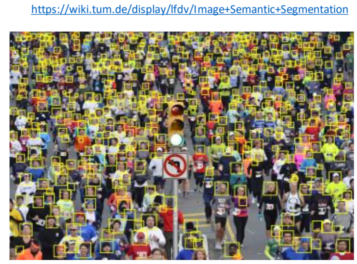
Introduction

Why to use transfer learning ?

Supervised learning methods need labeled data!

- It is hard to label a vast amount of data;
- It requires time, tools and people;
- It is expensive;
- After a single shot labeling, there is no guarantee that labels are right;
- Depending on the task it is more complex (like semantic segmentation or object detection).

https://wiki.tum.de/display/lfdv/Image+Semantic+Segmentation



T. T. Pham (2015)

Introduction

Why to use transfer learning ?

Supervised learning methods need labeled data!

Why not to get the labeled data from users?

- Because of privacy!
- It's just not possible to use the data from users without explicit permission;
- Big datasets got down by the issues involving the General Data Protection Regulation (GDPR);

Introduction

Why to use transfer learning ?

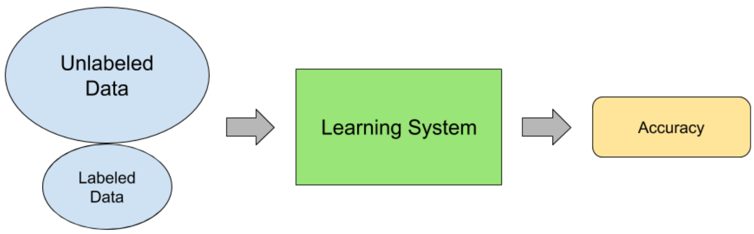
Supervised learning methods need labeled data!

Why not to get the labeled data from users?

- The GDPR is a European law on data protection that aims to give control to individuals over their personal data;
- It forces companies that collect data from users to say when, how and the purpose to collect;
- Users can request to remove their data from company database at any time.

Introduction

In semi-supervised learning it is possible to have a large amount of unlabeled data and a small set of labeled data, assuming that both data have the same distribution.

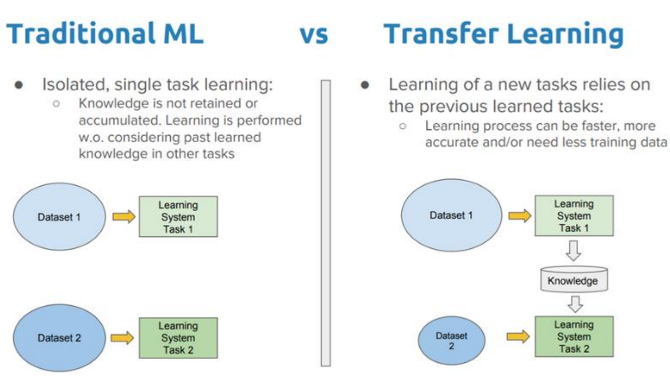


What's the minimum number of labeled samples to achieve a good performance?

Introduction

Traditional ML problems are designed to work in isolation. The transfer learning allow to transfer knowledge between different domains or tasks, reducing the isolation paradigm of having two different problems.

By using knowledge obtained from task 1 we can use less training data to achieve good performance on the task 2.



D. Sarkar (2018)

Introduction

Datasets

These are just a few examples that can be used for transfer learning.

IMAGENET
1.4 million images with 1k categories.
Object category classification and detection.

COCO Common Objects in Context
2.5 million labeled instances of objects.
328k images in 91 categories.

Open Images Dataset
9.2 million labeled images.
Classification, object detection and visual relationship detection.

Labeled Faces in the Wild

13k+ images of faces collected from the web.
Labeled with the name of the person.

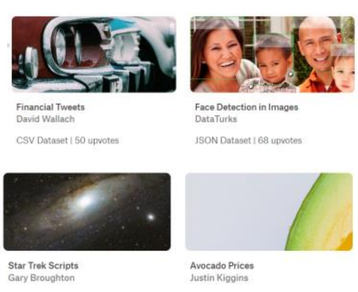
Introduction

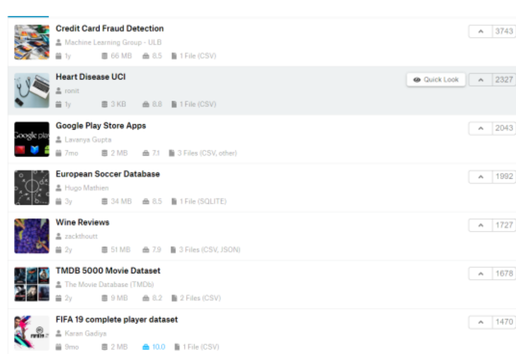
Datasets

The Kaggle website has a lot of public datasets and competitions.
They can be used to transfer learning to a similar task.

kaggle

www.kaggle.com





Introduction

Datasets

It’s common to find a dataset or a pre-trained model with the same domain of your problem but with different labels.

For example:

The ImageNet dataset has a lot of images labeled as electric guitar or acoustic guitar.

What if I want to train a model to recognize different types of guitars?

In this case, by using transfer learning a small amount of labeled data would be enough.

Introduction

If wrongly applied, transfer learning can led to a negative transfer. A decrease of the accuracy in the target model in comparison to not using transfer learning.

In this tutorial we try to answer three important questions:

- What to transfer?
- When to transfer?
- How to transfer?

Definition

We can split the definition of transfer learning into two high level concepts:

Domain $\mathcal{D}$

Most related to the distribution of the data to be used for training

Learning task $\mathcal{T}$

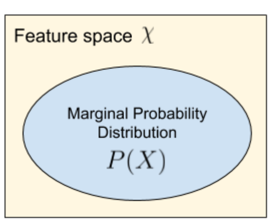
Given the domain $\mathcal{D}$, the task is defined by a label space and a predictive function

Pan and Yang, 2011

Definition

Domain $\mathcal{D}$

- A feature space $\mathcal{X}$
- A marginal probability distribution $P(X)$



- $\mathcal{X}$ is the entire space of possible features
- X is a particular learning sample with a subset of features from $\mathcal{X}$, where:

$$X = \{x_1, \ldots, x_n\} \in \mathcal{X}$$

- x_i is the i^{th} feature contained in some samples of the training data

Definition

Domain $\mathcal{D}$ – Example: Document classification

- $\mathcal{X}$ is the space of all term vectors
- x_i is the i^{th} term vector corresponding to some documents
- X is a particular learning sample

Feature Space = { "planet", "star", "galaxy", "how", "the", "and", "did", "first", … , "particles", "atoms", "molecules" }

X_1 = { "how", "did", "the", "first", "galaxy", "form" }

X_2 = { "how", "atoms", "and", "molecules", "Interact" }

Definition

Domain $\mathcal{D}$ – Example: Object classification

In a computer vision problem:

- $\mathcal{X}$ is the space of all **features** contained in all the images of the **dataset**
- x_i is the i^{th} **feature** present in some images
- X is a particular learning sample (the **features** contained in a **single image**)

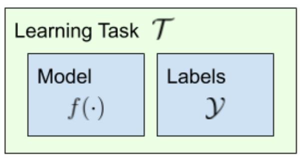
If two domains are different, they may have different **feature spaces** or different **marginal distributions.**

$X = \{x_i, \ldots, x_n\}$

Definition

Learning task $\mathcal{T}$

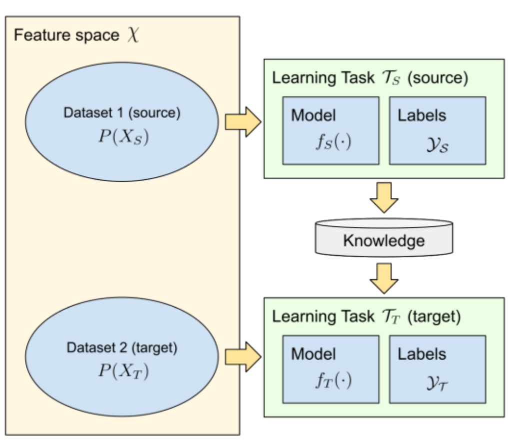
- A set of all possible labels $\mathcal{Y}$
- A model or predictive function $f(\cdot)$



- It can be **learned** from the training data, which consists of a collection of pairs $\{x_i, y_i\}$
- $x_i \in X$ and $y_i \in \mathcal{Y}$
- The function $f(\cdot)$ is used to predict the label y from an instance x

Definition

Transfer Learning



Having:

- a source domain $\mathcal{D}_s$
- a learning task $\mathcal{T}_s$
- a target domain $\mathcal{D}_T$
- a learning task $\mathcal{T}_T$

Definition:

Transfer Learning is a process that will help to improve the function $f_T(\cdot)$ in the target learning task based on the knowledge obtained from $\mathcal{D}_s$ and $\mathcal{T}_s$, where $\mathcal{D}_s \neq \mathcal{D}_T$, or $\mathcal{T}_s \neq \mathcal{T}_T$.

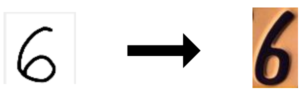
Definition

Transfer Learning

For source and target **domains** to be different:

$\boldsymbol{\mathcal{X}_s} \neq \boldsymbol{\mathcal{X}_T}$ → The feature space of the source and target are different, e.g. the documents that are written in two different languages.

$\boldsymbol{P(\mathcal{X}_s)} \neq \boldsymbol{P(\mathcal{X}_T)}$ → The marginal distributions of source and target domain are different, e.g. the documents that discuss different topics. Generally known as domain adaptation.



Wulfmeier, Bewley and Posner (2017)

Definition

Transfer Learning

For source and target **tasks** to be different:

$\mathcal{Y}_{\mathrm{S}} \neq \mathcal{Y}_{\mathrm{T}}$ → The label spaces between the two tasks are different, e.g. documents need to be assigned different labels in the target task

Is Receipt ? → { receipt, invoice, letter, email, report }

$f_S(\cdot) \neq f_T(\cdot)$ → The model or predictive function of the source and target tasks are different.

Transfer Learning Strategies

Three important questions:

- ❑ *What to transfer?*
  - Understand which part of knowledge from source can be used to improve the performance on the target, it's important to identify which part of the knowledge is specific or not.
- ❑ *When to transfer?*
  - The decision of transfer or not. In which situation source domain and target domain are not enough related or won't help to improve the performance on the target, resulting in negative transfer.
- ❑ *How to transfer?*
  - The algorithms that needs to be develop to transfer the knowledge will answer this question.

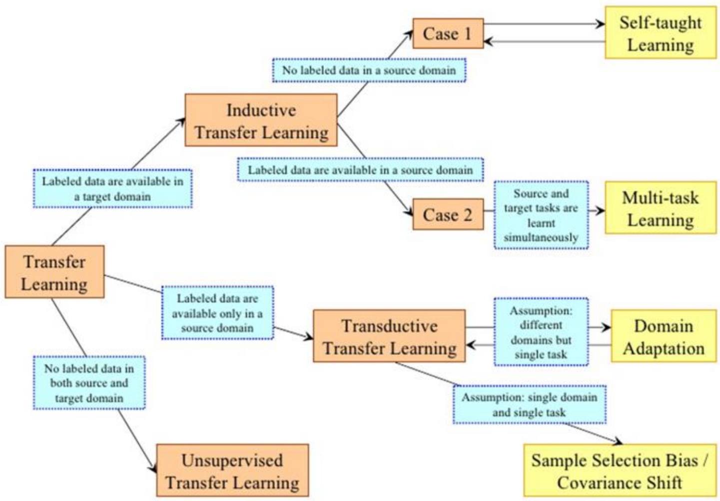
Transfer Learning Strategies

Based on the definition and on different existing scenarios between source and target, we can categorize transfer learning under three strategies/settings:

- ➢ Inductive Transfer Learning
- ➢ Transductive Transfer Learning
- ➢ Unsupervised Transfer Learning

Some authors call this **strategies** as **settings**.

Transfer Learning Strategies



Pan and Yang (2009)

Transfer Learning Strategies

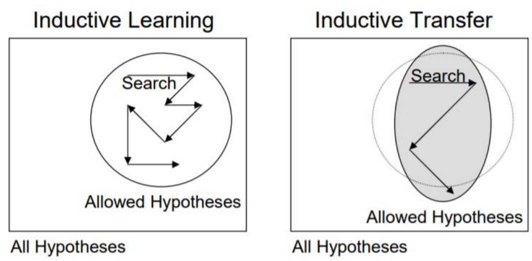
Inductive Transfer Learning

- Refers to use cases where source and target tasks are different and domains can be related or not.
- Some labeled data in the target is needed to induce the target objective predictive function $f_T(\cdot)$.
- Using the inductive biases from the source domain this setting tries to improve the task in the target.
- Commonly used in deep learning applications.

Transfer Learning Strategies

Inductive Transfer Learning

The target task inductive bias is chosen or adjusted based on the source task knowledge. It can be viewed as a directed search through a specified hypothesis space.



L. Torrey & J. Shavlik (2009)

Transfer Learning Strategies

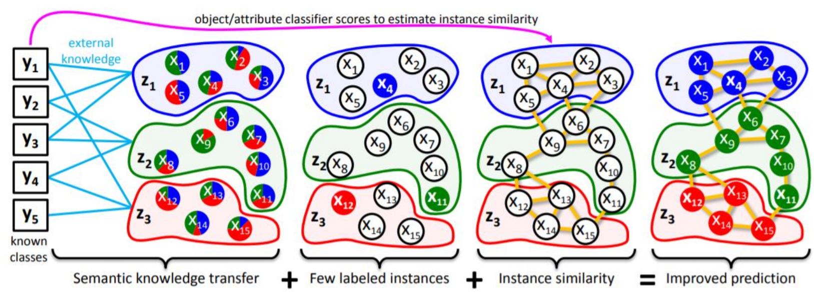
Transductive Transfer Learning

- The source and target tasks are similar, while the source and target domains are different.
- There is no <u>labeled</u> data available in the target, only in the source, which is similar to semi-supervised learning.
- Some <u>unlabeled</u> target domain data must be available at training time.

Transfer Learning Strategies

Transductive Transfer Learning

Example of a transductive transfer learning from "Transfer Learning in a Transductive Setting", Rohrbach, Ebert and Schiele (2013).



Transfer Learning Strategies

Unsupervised Transfer Learning

- Similar to inductive transfer learning.
- Tasks are different between the target and the source.
- Focus on resolve unsupervised learning tasks in the target domain, like clustering and dimensionality reduction.

Transfer Learning Approaches

Using the three described strategies/settings, there are different approaches that can be applied in terms of "What to transfer" (Pan and Yang, 2009):

- Instance Transfer
- Feature Representation Transfer
- Parameter Transfer
- Relational Knowledge Transfer

Transfer Learning Approaches

Instance Transfer

- Although the source domain data cannot be reused directly, there are certain parts of the data that can still be reused together with a few labeled data in the target domain.
- In deep learning applications this process can be called as re-weighting or fine-tuning.

Transfer Learning Approaches

Feature Representation Transfer

- The knowledge is encoded into a feature representation, which is learned from the source and used to improve the target task.
- It's expected that the data in the source domain to be enough to generate a good feature representation for the target.
- Example: Use of models pre-trained with the ImageNet dataset as feature extractor.
- Substantial gains over hand engineered features (SIFT, SURF, ORB, etc.) when transferred to other tasks.

Transfer Learning Approaches

Parameter Transfer

- Source and target tasks shares some parameters or prior distribution of hyper-parameters of the models.
- The transferred knowledge is, in fact, the shared parameters between the models.
- In parameter transfer, there are two models and some of the parameters may be different.
- Different of multi-task learning, the source and target tasks are not learned simultaneously.

Transfer Learning Approaches

Relational Knowledge Transfer

- Deals with the transfer of knowledge between relational domains.
- Assumption that the data in the source and target contains similar relationship and the knowledge to be transferred is this relationship among the data.
- Example: Models that use social network data.

Transfer Learning Approaches

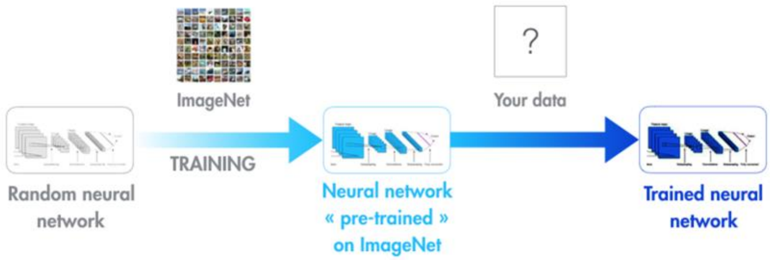
The following table shows the different approches used in different settings/strategies:

	Inductive Transfer Learning	Transductive Transfer Learning	Unsupervised Transfer Learning
Instance-transfer	X	X	
Feature-representation-transfer	X	X	X
Parameter-transfer	X		
Relational-knowledge-transfer	X		

Pan and Yang (2009)

Transfer Learning Applied to Deep Learning

With the advances in the field of deep learning, research related to transfer learning has focused on this kind of neural network architecture.



Transfer Learning Applied to Deep Learning

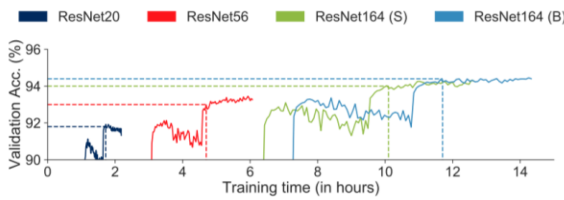
C. Tan et al. (2018) in "A Survey on Deep Transfer Learning" categorize transfer learning into four approaches:

- *Instance-based transfer*, which is the same idea that we described before.
- *Mapping-based transfer*, that uses data from both source and target domains to create a new data space for training.
- *Network-based transfer,* the same that we described before for the feature representation transfer.
- *Adversarial-based transfer*, which applies a technology inspired by the generative adversarial networks (GANs) to find generative features suitable for source and target domains.

Transfer Learning Applied to Deep Learning

Instead of training a deep learning model from scratch with a lot of data and taking days of training, it's possible to get a model trained on a different domain for a different source task and adapt it to the desired target task.

NVIDIA K80 GPU
Compute Capability = 3.7



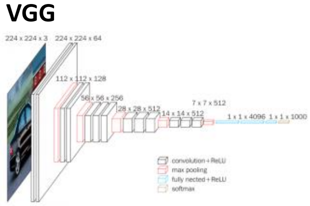
Accuracy threshold	Time to threshold	Fastest model
91.8%	1h 43m	ResNet20
93%	4h 42m	ResNet56
94%	10h 6m	ResNet164 (simple)
94.4%	11h 42m	ResNet164 (bottleneck)

DAWNBench - Coleman et al. (NIPS2017)

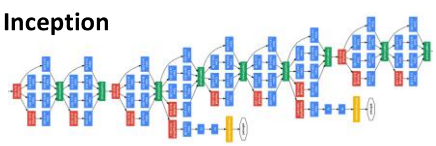
Transfer Learning Applied to Deep Learning

There are many deep learning architectures with pre-trained models that have achieved good performance within different tasks and can be leveraged for transfer learning to related tasks.

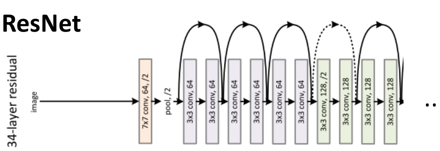
VGG



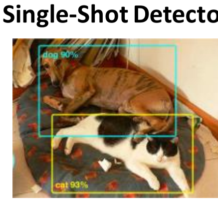
Inception



ResNet



Single-Shot Detector



Transfer Learning Applied to Deep Learning

Official models available in Keras pre-trained in ImageNet

- Xception
- VGG16
- VGG19
- ResNet, ResNetV2
- InceptionV3
- InceptionResNetV2
- MobileNet
- MobileNetV2
- DenseNet
- NASNet

https://keras.io/applications/

Transfer Learning Applied to Deep Learning

Documentation for individual models

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	-
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
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	-

The top-1 and top-5 accuracy refers to the model's performance on the ImageNet validation dataset.

https://keras.io/applications/

Transfer Learning Applied to Deep Learning

TensorFlow Object Recognition Model Zoo

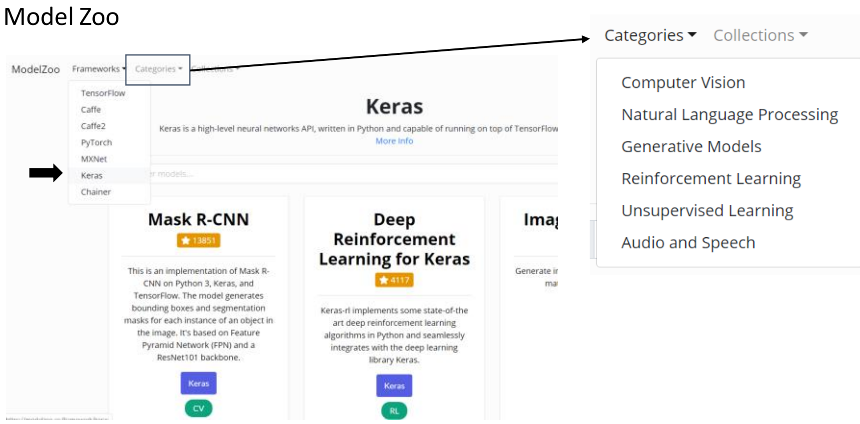
https://github.com/tensorflow/models

COCO-trained models

Model name	Speed (ms)	COCO mAP[^1]	Outputs
ssd_mobilenet_v1_coco	30	21	Boxes
ssd_mobilenet_v1_0.75_depth_coco ☆	26	18	Boxes
ssd_mobilenet_v1_quantized_coco ☆	29	18	Boxes
ssd_mobilenet_v1_0.75_depth_quantized_coco ☆	29	16	Boxes
ssd_mobilenet_v1_ppn_coco ☆	26	20	Boxes
ssd_mobilenet_v1_fpn_coco ☆	56	32	Boxes
ssd_resnet_50_fpn_coco ☆	76	35	Boxes
ssd_mobilenet_v2_coco	31	22	Boxes
ssd_mobilenet_v2_quantized_coco	29	22	Boxes
ssdlite_mobilenet_v2_coco	27	22	Boxes
ssd_inception_v2_coco	42	24	Boxes
faster_rcnn_inception_v2_coco	58	28	Boxes
faster_rcnn_resnet50_coco	89	30	Boxes
faster_rcnn_resnet50_lowproposals_coco	64		Boxes
rfcn_resnet101_coco	92	30	Boxes
faster_rcnn_resnet101_coco	106	32	Boxes
faster_rcnn_resnet101_lowproposals_coco	82		Boxes

Transfer Learning Applied to Deep Learning

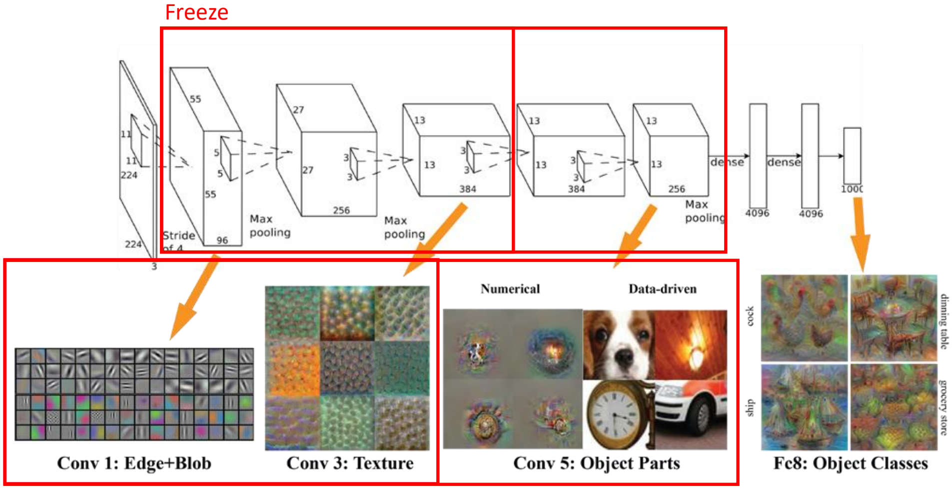
Model Zoo



https://modelzoo.co/

Transfer Learning Applied to Deep Learning

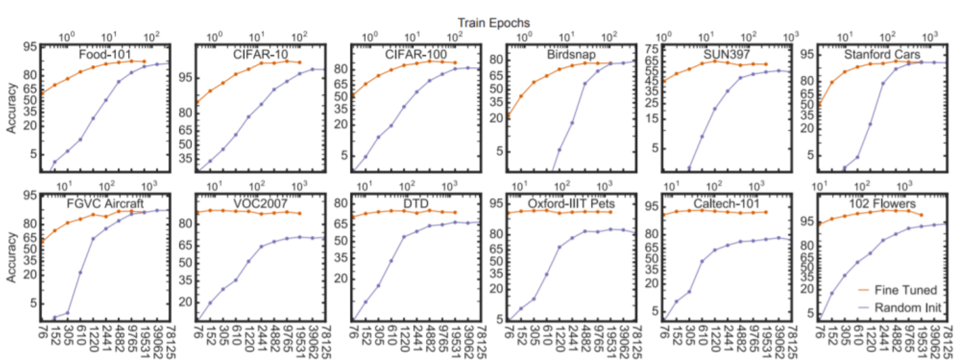
Feature Representation Transfer



https://www.cc.gatech.edu/~hays/compvision2017/proj6/

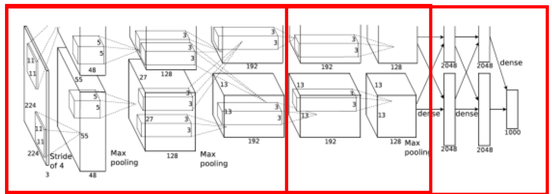
Transfer Learning Applied to Deep Learning

Do Better ImageNet Models Transfer Better?
https://arxiv.org/abs/1805.08974



Transfer Learning Applied to Deep Learning

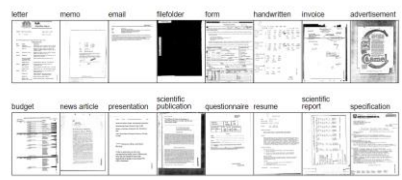
To decide between fine-tuning or freeze, it's important to consider the distribution of data.



ImageNet
http://www.image-net.org/

VGG Flowers
http://www.robots.ox.ac.uk/~vgg/data/flowers/17/index.html

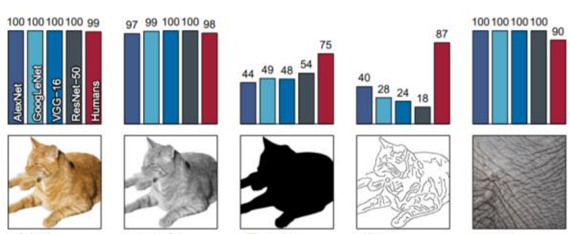
RVL-CDIP
https://www.cs.cmu.edu/~aharley/rvl-cdip/

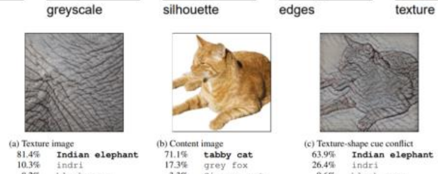


Transfer Learning Applied to Deep Learning

A recent research shown that the CNNs pre-trained with ImageNet dataset are biased towards texture, causing negative transfer depending on the use case.

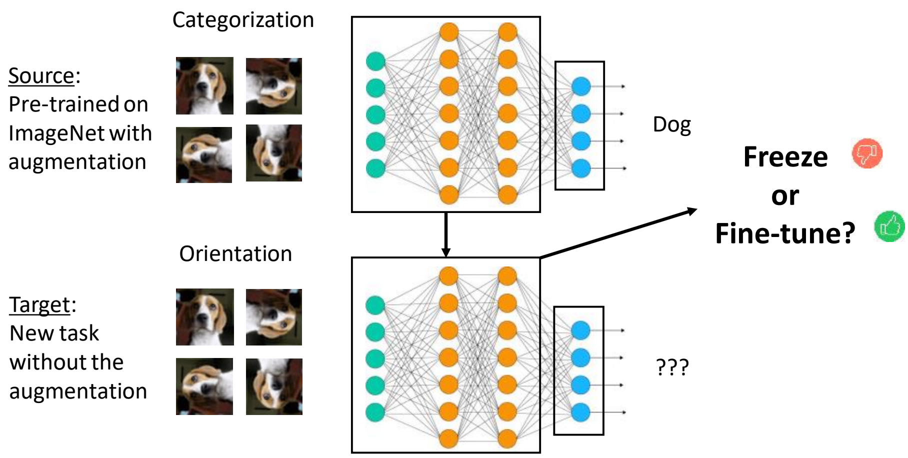
ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness
https://arxiv.org/abs/1811.12231





Transfer Learning Applied to Deep Learning

Available pre-trained models may have been trained with augmentation. If the target task needs to recognize patterns without the augmentation applied in the source domain, some or all layers needs to be fine-tuned.



Types of Transfer Learning

The research of transfer learning appears in the literature with different names, sometimes focused on different techniques and sometimes just with a different name.

- Domain Adaptation
- Domain Confusion
- Multitask Learning
- One-Shot Learning
- Zero-Shot Learning
- Meta Learning

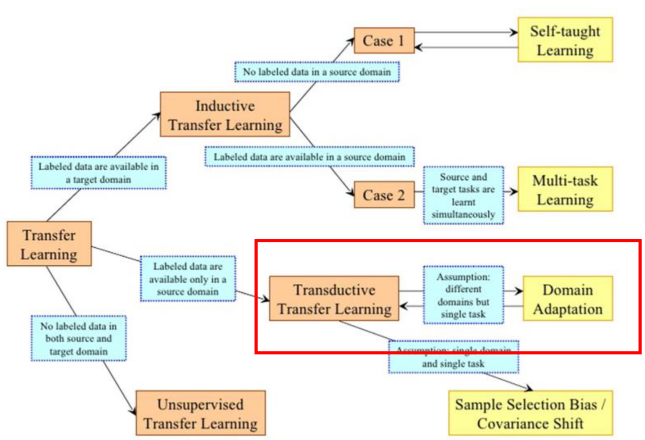
Types of Transfer Learning

Domain Adaptation

- Feature spaces are similar between source and target.
- But marginal probability distributions are different.
- There is a shift between both domains required to implement an adaptation technique prior to use transfer learning.

Types of Transfer Learning

Domain Adaptation



Pan and Yang (2009)

Types of Transfer Learning

Domain Adaptation

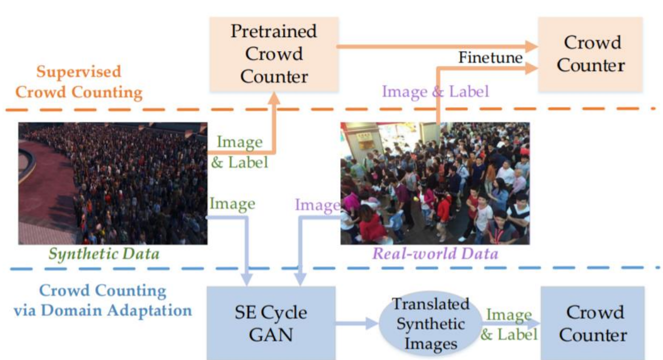
<u>Examples</u>:

- ➢ Spam Filtering - Adapting a model from one user (the source distribution) to a new one who receives significantly different emails (the target distribution).
- ➢ Sentiment Analysis - A corpus of movie reviews labeled as positive or negative would be different from a corpus of product-review sentiments. A classifier trained on movie-review sentiment would see a different distribution if utilized to classify product reviews.

Types of Transfer Learning

Domain Adaptation

CV Example: Learning from Synthetic Data for Crowd Counting in the Wild
https://arxiv.org/pdf/1903.03303.pdf



Types of Transfer Learning

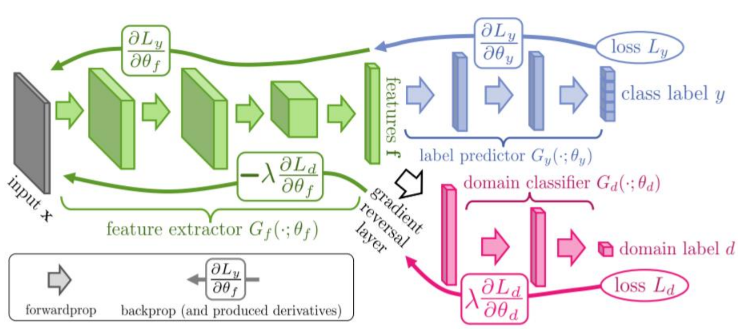
Domain Confusion

- This technique takes advantage of the feature representation learned by deep learning models to learn domain-invariant features and improve transferability across domains.
- Push learned representations in both domains to be as similar as possible.
- Embed domain adaptation into the process of learning representation, so that the final classification decisions are made based on features that are both discriminative and invariant to the change of domains;
- Add another objective to the source model to encourage similarity by confusing the domain itself, hence domain confusion.

Types of Transfer Learning

Domain Confusion

Domain-Adversarial Training of Neural Networks
https://arxiv.org/abs/1505.07818



Types of Transfer Learning

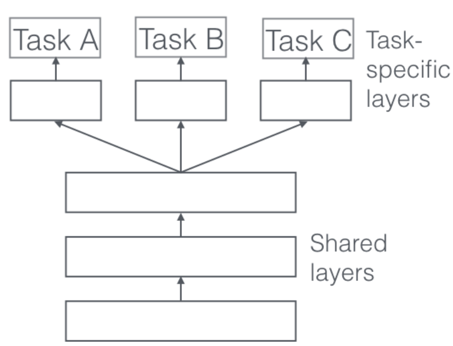
Multitask Learning

- Multitask learning aim to leverage useful information contained in multiple related tasks to help improve the generalization performance of all the tasks.

- This is a different type of transfer learning, where both source and target tasks are learned at the same time, taking advantage of the common features existing in both domains to obtain the best balance in performance for both tasks.

Types of Transfer Learning

Multitask Learning



https://towardsdatascience.com/multitask-learning-teach-your-ai-more-to-make-it-better-dde116c2cd40

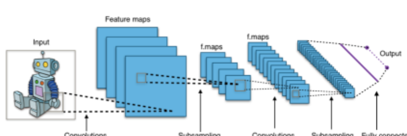
Types of Transfer Learning

One-Shot Learning

- A person can easily identify an object after having seen it for the first time, but not the existing machines. For instance, once it is shown to a child what an applelooks like, they can easily identify a different variety of apple (with one or a few training examples);
- One-shot learning is a variant of transfer learning, that tries to infer the required output based on just one <u>or a few</u> training examples.
- This is helpful in real-world scenarios where it is not possible to have labeled data for every possible class (if it is a classification task), and in scenarios where new classes can be added often.

Types of Transfer Learning

One-Shot/Few-Shot Learning



100 categories

Train, Val, Test
99% Accuracy

What is the one-shot accuracy for more categories?

100 10

+10 categories

If improving generalization for these 10 classes, can we easily improve for more 10?

100 10 10

+10 categories

Types of Transfer Learning

Zero-Shot Learning

- This technique aim to intelligently apply previously learned knowledge to help future recognition tasks. It relies on no labeled examples to learn a task and can also be referred as translated learning.
- Learn the relationship between input data and the semantic features, then use the learned relationship in a two step prediction procedure to recover the class label for novel input data;
- It will work for input data from a novel class if that class is included in the semantic knowledge base;
- For example, presenting a picture of a zebra to the system that have never seen a zebra but has seen a horse and also taught that a zebra looks like a horse but with stripes.

Types of Transfer Learning

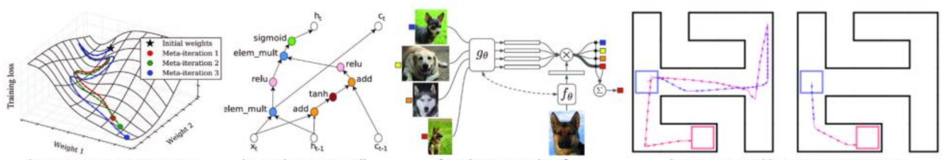
Meta-Learning

- The idea was first described by J. Schmidhuber as *learning to learn* in the late 1980s and Y. Bengio et al. in early 1990s.
- It has also been studied with modern approaches, for neural networks optimization, to find better network architectures, in few-shot image recognition (related to *one-shot learning and zero-shot learning*) and fast reinforcement learning.
- These systems are trained with different variety of tasks such as they can learn new tasks faster and with few shot examples, based only on prior knowledge on how to resolve these tasks.

Types of Transfer Learning

Meta-Learning

Recently meta-learning has become a hot topic, with a flurry of recent papers.



hyperparameter optimization Maclaurin et al.'15 | learned recurrent cell Zoph & Le'17 | few-shot image classifier Vinyals et al.'16 | learning to quickly navigate new mazes Duan et al.'16

https://bair.berkeley.edu/blog/2017/07/18/learning-to-learn/

Conclusion and Future of Research

Transfer Learning Advantages

- Typically transfer learning enables us to build more robust models which can perform a wide variety of tasks.
- Helps solve complex real-world problems when having little or almost no labeled data availability;
- Ease of transferring knowledge from one model to another based on domains and tasks;

Conclusion and Future of Research

Transfer Learning Challenges

- An open question in the field of transfer learning is related to avoiding *negative transfer*, where the transfer of knowledge from the source to the target does not lead to any improvement.

- Avoiding negative transfer is crucial to implement transfer learning.

- There can be various reasons for negative transfer, such as cases when the source task is not sufficiently related to the target task, or if the transfer method could not leverage the relationship between the source and target tasks very well.

- There is a need to study transferability between the source and the target.

Conclusion and Future of Research

Future Scope

- One new technique that has gained attention is called Generative Adversarial Networks (GANs), which is considered the most interesting idea in the last 10 years in ML by most researchers. Considering the potential of this idea and the impressive results published in many papers, there is also an opportunity for improvement using transfer learning with these models.

- It can be useful to transfer knowledge learned by the generators or discriminators across different domains, and to make the training faster.

Conclusion and Future of Research

Future Scope

- Methods of transfer learning that help to train models with small amount of data are very promising areas of research, like *one-shot learning, zero-shot learning,* and *meta-learning*.

- Quantifying the transfer and measure relatedness between tasks.

Thank you!

Ricardo Ribani
rribani@hotmail.com

Mauricio Marengoni
mmarengoni@Mackenzie.br

Sponsorship: Mack Pesquisa