Deep Learning & Engineering Applications

7. Transfer Learning

JIDONG J. YANG

COLLEGE OF ENGINEERING

UNIVERSITY OF GEORGIA

Transfer Learning

Learning how to play soccer → Learn how to play basketball

Learning how to play tennis → Learn how to play table tennis

Learning linear algebra & statistics → Learn deep learning

Learning in classroom → Learning in workplace

Basically, apply knowledge gained while solving one problem to solve a different but related problem.

Motivated by the concept of "Learning to Learn."
- Focus on the need for lifelong machine-learning methods that retain and reuse previously learned knowledge

Formal Definition

A domain, D, is defined as $D = \{X, P(X)\}$

A task, $T = \{y, f(.)\}$ or $T = \{y, P(y|X\}$

Transfer Learning

- Given:
  - A source domain D_s and learning task T_s
  - A target domain D_T and learning task T_T
- Improve learning of the T_T in D_T using the knowledge in D_s and T_s

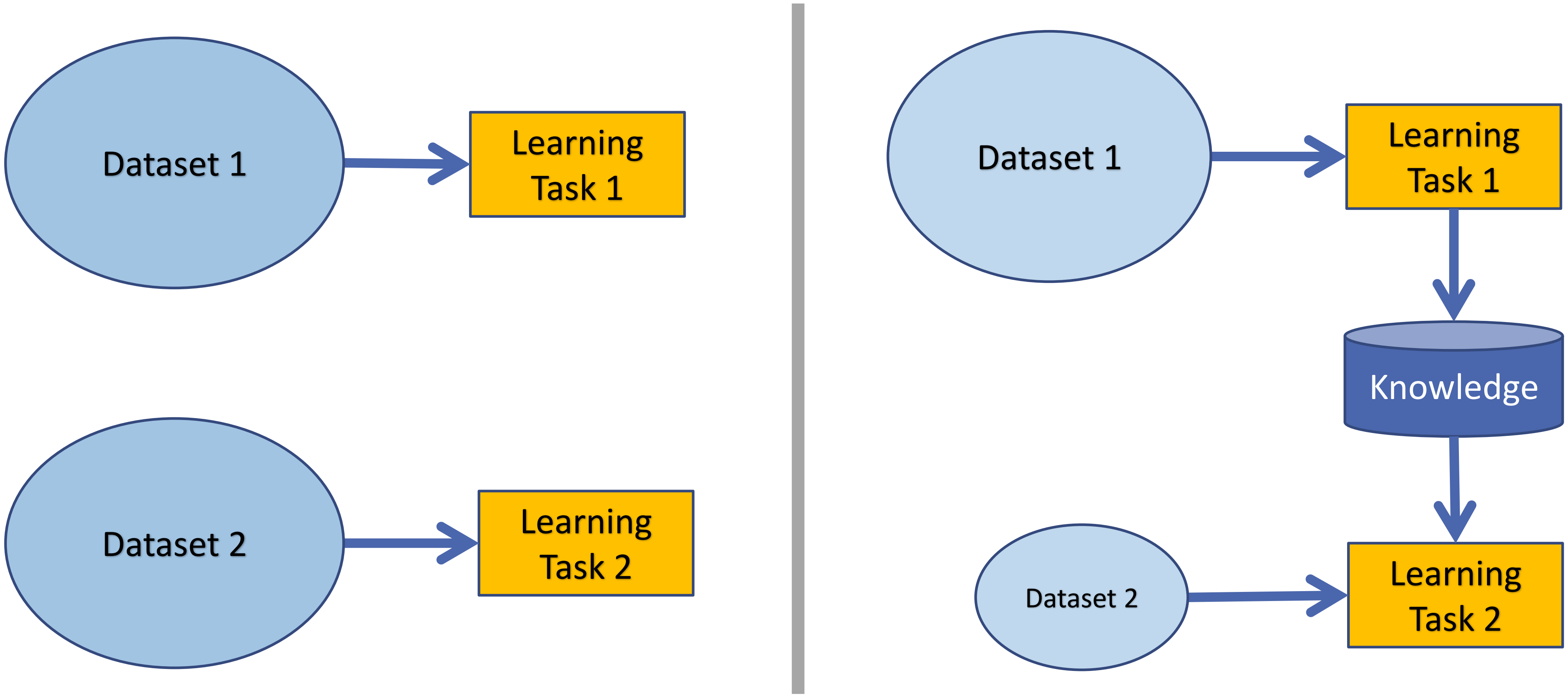
[Pan and Yang, 2009] https://www.cse.ust.hk/~qyang/Docs/2009/tkde_transfer_learning.pdf

Relationship between Traditional ML and Various Transfer Learning Settings

Setting		D_s and D_T	T_s and T_T
Traditional ML		Same	Same
Transfer Learning	Inductive	Same	Different but related
	Transductive	Different but related	Same
	Unsupervised	Different but related	Different but related

[Pan and Yang, 2009] https://www.cse.ust.hk/~qyang/Docs/2009/tkde_transfer_learning.pdf

Traditional ML vs Transfer Learning



Energy and Policy Considerations for Deep Learning

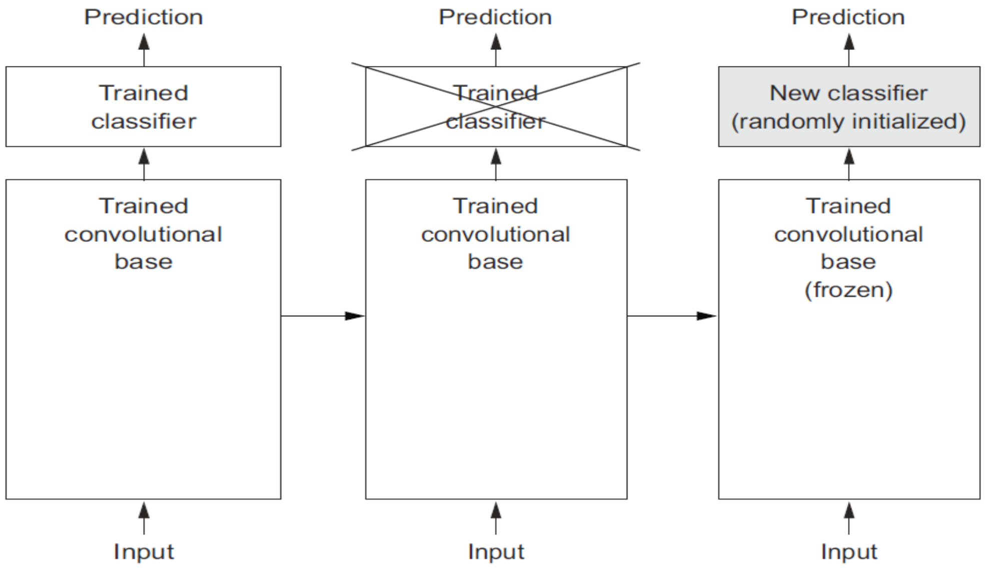
Model	Hardware	Power (W)	Hours	kWh·PUE	CO_2e	Cloud compute cost
$\text{Transformer}_{base}$	P100x8	1415.78	12	27	26	$41–$140
Transformer_{big}	P100x8	1515.43	84	201	192	$289–$981
ELMo	P100x3	517.66	336	275	262	$433–$1472
BERT_{base}	V100x64	12,041.51	79	1507	1438	$3751–$12,571
BERT_{base}	TPUv2x16	—	96	—	—	$2074–$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	$942,973–$3,201,722
NAS	TPUv2x1	—	32,623	—	—	$44,055–$146,848
GPT-2	TPUv3x32	—	168	—	—	$12,902–$43,008

Table 3: Estimated cost of training a model in terms of CO_2 emissions (lbs) and cloud compute cost (USD).[7] Power and carbon footprint are omitted for TPUs due to lack of public information on power draw for this hardware.

[Strubell et al., 2019] https://arxiv.org/pdf/1906.02243.pdf

Transfer Learning with CNN

- A CNN model usually includes two parts:
  - Convolutional base network
  - Classifier
- Transfer Learning:
  - Take the base network of a previously trained network.
  - Train a new fully connected classifier on top of the output of the base network.
  - Finetune the entire network.



Example:
FCN with ResNet-18

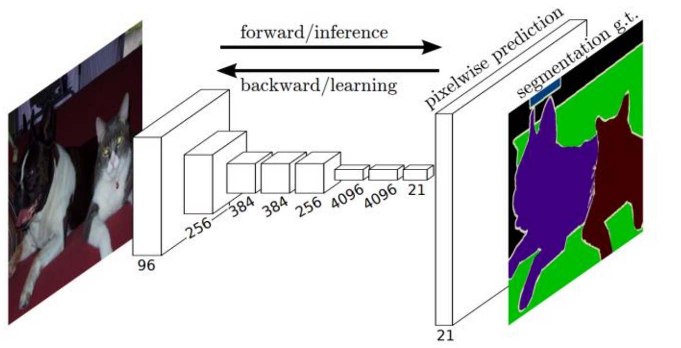


Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.