Discussion of Cepelewicz: “Where we see shapes AI sees textures”

For the AI in Medical Imaging and Signal Processing Journal Club

André Carrington PhD PEng CISSP

January 28, 2020

The Ottawa Hospital | L’Hôpital d’Ottawa
RESEARCH INSTITUTE | INSTITUT DE RECHERCHE

Department of Medicine

uOttawa

Context: classifying images with deep learning (DL)

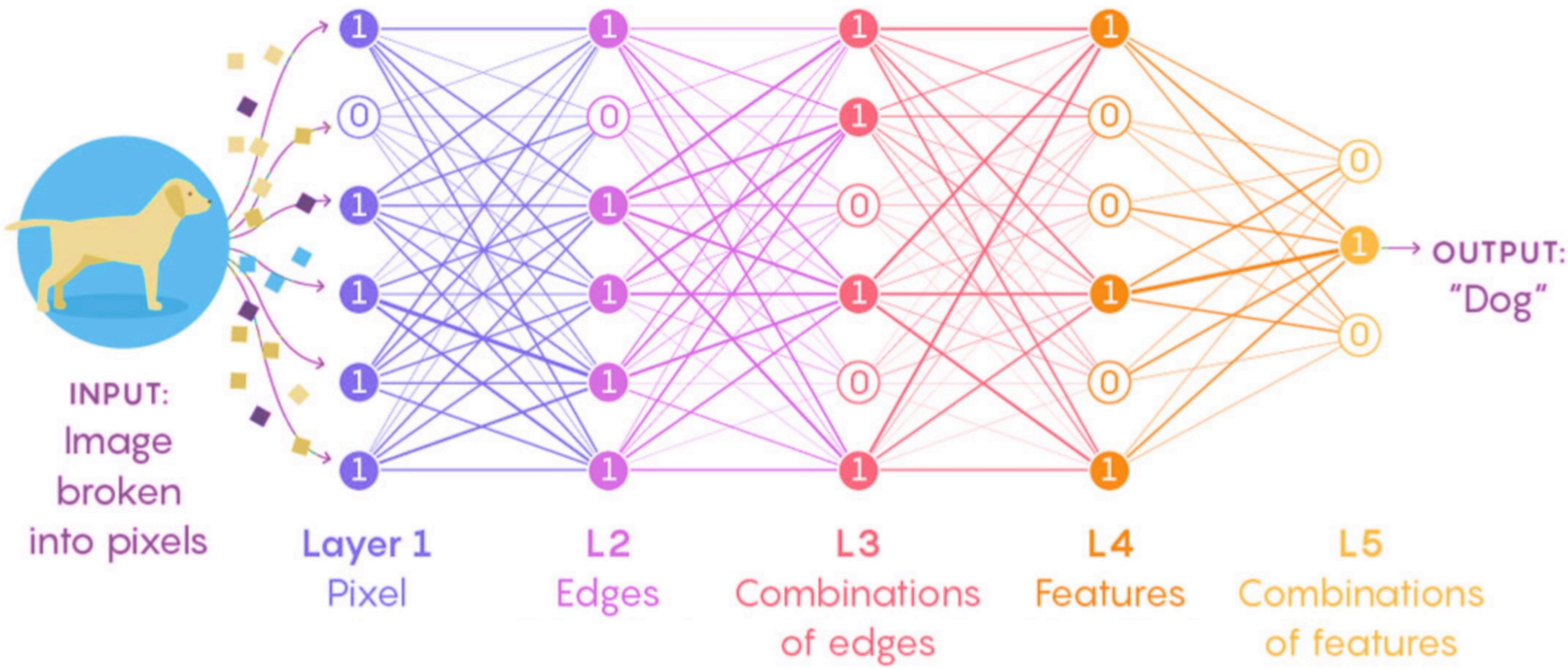


Figure adapted from Quantum Magazine

In reality, the features (information), formed at each layer is more mysterious…

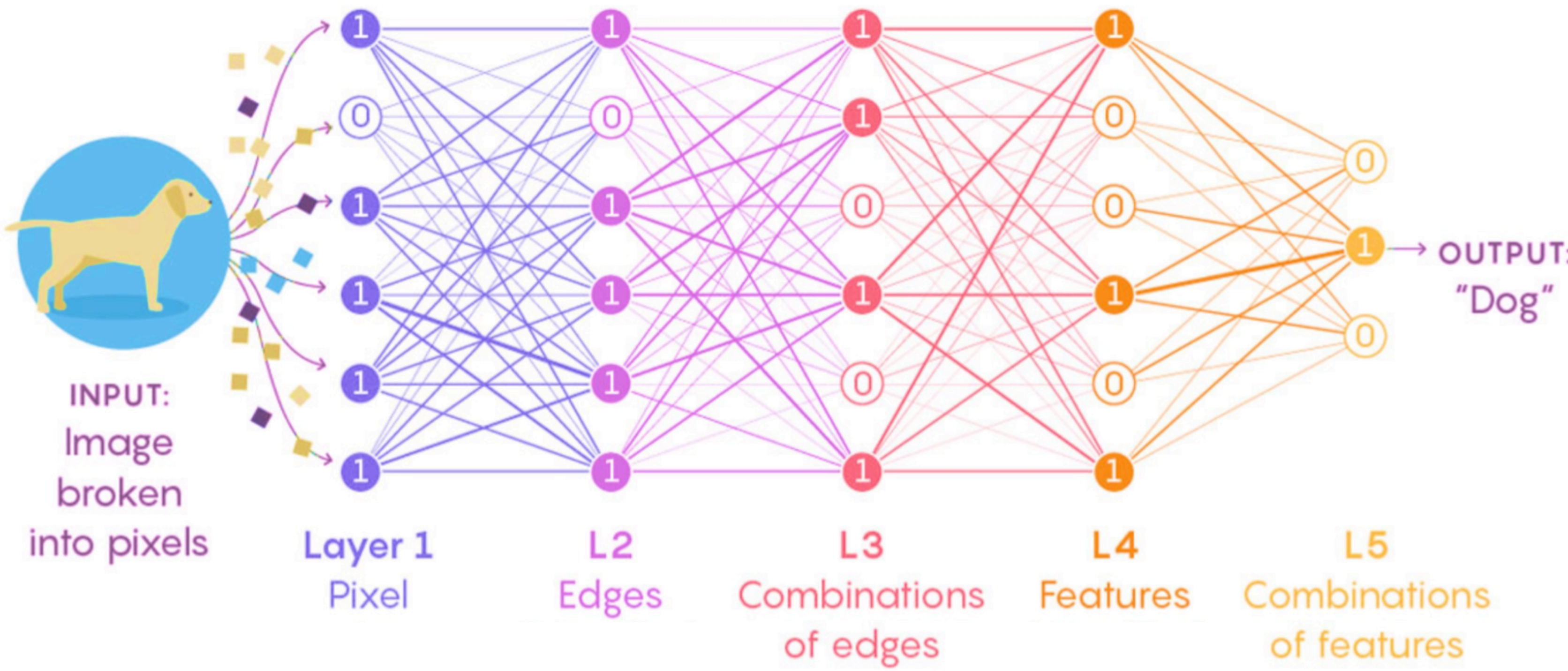


Figure adapted from Quantum Magazine

The Ottawa Hospital RESEARCH INSTITUTE | L'Hôpital d'Ottawa INSTITUT DE RECHERCHE

Department of Medicine

uOttawa

Central theme/claim

- *"To researchers' surprise, deep learning vision algorithms **often fail** at classifying images because they mostly take cues from textures, not shapes."*

The Ottawa Hospital | L'Hôpital d'Ottawa
RESEARCH INSTITUTE | INSTITUT DE RECHERCHE

Department of Medicine

uOttawa

Central theme/claim

- *"To researchers' surprise, deep learning vision algorithms **often fail** at classifying images because they mostly take cues from textures, not shapes."*

    - DL **often succeeds** (internal validation)*

Central theme/claim

- *"To researchers' surprise, deep learning vision algorithms **often fail** at classifying images because they mostly take cues from textures, not shapes."*

  - DL **often succeeds** (internal validation)*
  - But **fails** to generalize (ext. validation)*
    - other machines, environments, cases; adversarial inputs

Experiment: painting cats with elephant skin

- *The classifier identified an elephant (by texture)*
- *Humans identified a cat (by shape)*

Experiment: painting cats with elephant skin

- *The classifier identified an elephant (by texture)*
- *Humans identified a cat (by shape)*
    - This is an adversarial example which may not be realistic for all domains, e.g., surgery*

Experiment: painting cats with elephant skin

- *The classifier identified an elephant (by texture)*
- *Humans identified a cat (by shape)*
  - This is an adversarial example which may not be realistic for all domains, e.g., surgery*
    - Mimics, obstructions and noise are different. Obstructions confuse shape.*
    - i.e., which feature trumps? who is right?

This raises a bigger question*

- Do we want computers to:
  - Think like us?*
  - Or differently (to compliment our thinking)?*

This raises a bigger question*

- Do we want computers to:
  - Think like us?*
  - Or differently (to compliment our thinking)?*
    - It depends on the application/objective*
    - It can be useful or ideal to have votes (or probabilities of class membership) from:*
      - a shape classifier* **and**
      - a texture classifier*

Experiment: making DL use shapes

- Paint irrelevant textures (on objects, background)
- Performance improved
- But the classifier could still be fooled with trivial changes

Examples of how image classification can fail*

- A boat identified because of water
- A horse identified because of a shifted trademark
- A criminal identified because of whitespace
- Or in other ways which are not easily explained

Bigger picture again*

- Classifiers usually do not know which information is ***supposed to be*** relevant*
    - i.e., they lack prior knowledge (Bayesian priors)*

Bigger picture again*

- Classifiers usually do not know which information is ***supposed to be*** relevant*
  - i.e., they lack prior knowledge (Bayesian priors)*
  - But parametric statistical methods do!*

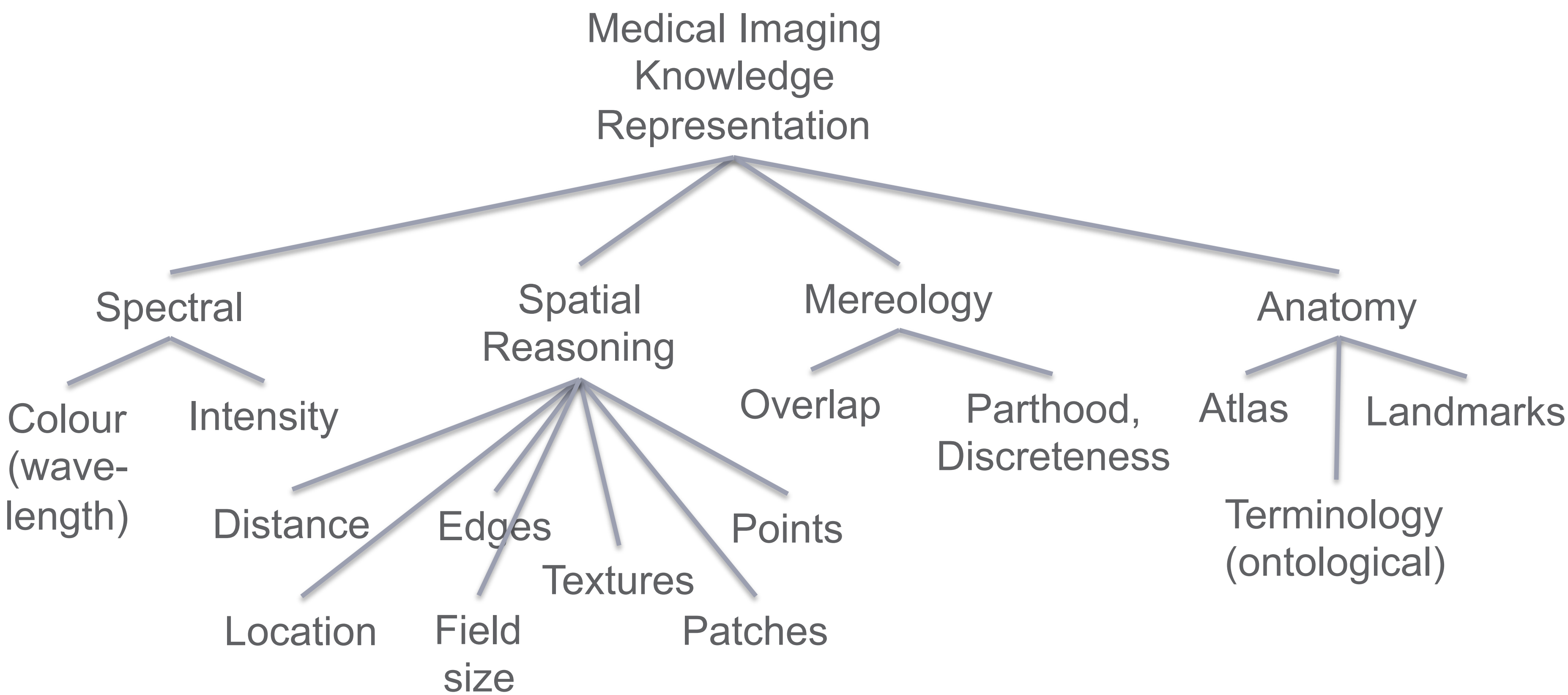
Bigger picture again*

- Classifiers usually do not know which information is ***supposed to be*** relevant*
    - i.e., they lack prior knowledge (Bayesian priors)*
    - But parametric statistical methods do!*
    - Manual feature engineering does!*
    - Knowledge bases do!*
    - Human-in-the-loop learning does!*

Bigger picture again*

- Classifiers usually do not know which information is ***supposed to be*** relevant*
    - i.e., they lack prior knowledge (Bayesian priors)*
    - But parametric statistical methods do!*
    - Manual feature engineering does!*
    - Knowledge bases do!*
    - Human-in-the-loop learning does!*
    - DL automates feature engineering*

A quick draft of key concepts*



The Ottawa Hospital | L'Hôpital d'Ottawa
RESEARCH INSTITUTE | INSTITUT DE RECHERCHE

Department of Medicine

uOttawa

Markov kernels*

$$\begin{array}{ccc} & -1 & \\ -1 & \boxed{4.1} & -1 \\ & -1 & \end{array}$$

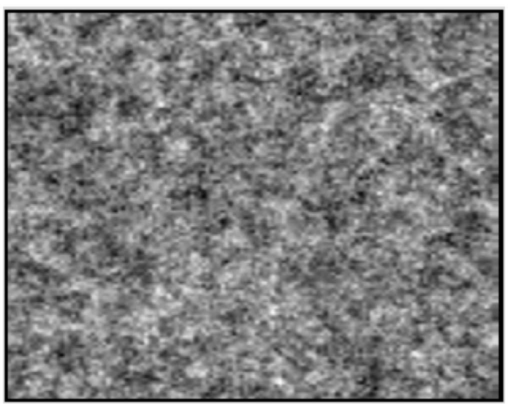
Membrane

$$\begin{array}{ccccc} & & 1 & & \\ & 2 & -8 & 2 & \\ 1 & -8 & \boxed{20.1} & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{array}$$

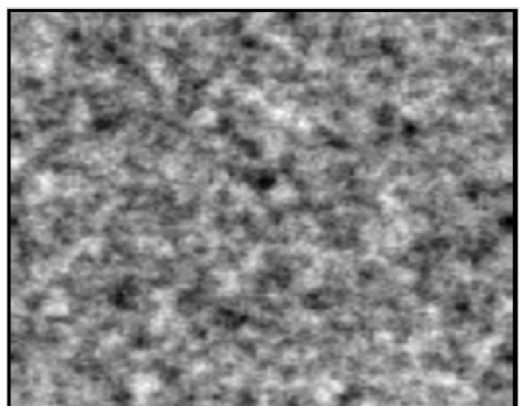
Thin-Plate

$$\begin{array}{ccccc} & -91 & 517 & 8 & \\ 58 & 1405 & -5508 & 1164 & 85 \\ -139 & -2498 & \boxed{10000} & -2498 & -139 \\ 85 & 1164 & -5508 & 1405 & 58 \\ & 8 & 517 & -91 & \end{array}$$

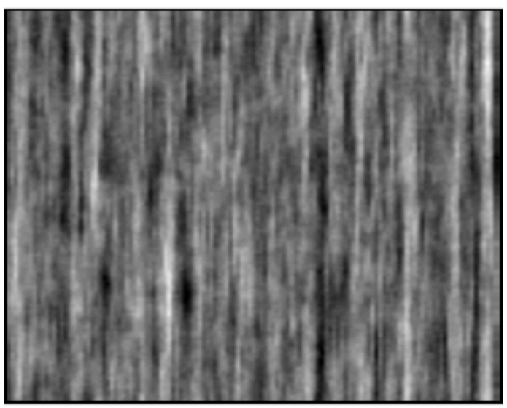
“Tree-Bark” [195]



1st order
3x3



3rd order
5x5



4th order
5x5

Figure adapted from Paul Fieguth’s Statistical Image Processing and Multidimensional Modeling, p.190.

Questions?

André Carrington PhD PEng CISSP

The Ottawa Hospital RESEARCH INSTITUTE | L'Hôpital d'Ottawa INSTITUT DE RECHERCHE

Department of Medicine

uOttawa