Neural Networks and Deep Learning: Introduction & Context

Nicolas Thome

Conservatoire National des Arts et Métiers (Cnam)
Département Informatique

Context: Big Data

- Superabundance of data: images, videos, audio, text, user traces, *etc*

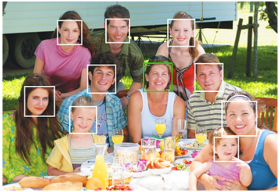
BBC: 2.4M videos

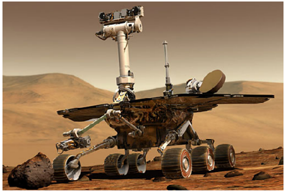
Social media,
e.g. Facebook: 1B each day

100M monitoring cameras

- Obvious need to access, search, or classify these data: **Recognition**

Recognition & Big Data





- Huge number of applications: mobile visual search, robotics, autonomous driving, augmented reality, medical imaging *etc*
- Leading track in major ML/CV conferences during the last decade

Focus on Visual Recognition: Perceiving Visual World

- Archetype of low-level signal
- Early 80's: master class problem
- Most impacted topic by deep learning

Focus on Visual Recognition: Perceiving Visual World

- Scene categorization: beach, moutain, city, *etc*
- Object localization: people, church, *etc*

Focus on Visual Recognition: Perceiving Visual World

- Context & attribute prediction: urban, outdoor, sunny, open, *etc*

Focus on Visual Recognition: Perceiving Visual World

- 3D layout, depth ordering: "mountain behind city"

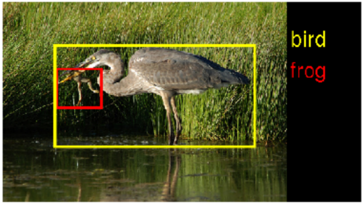
Focus on Visual Recognition: Perceiving Visual World

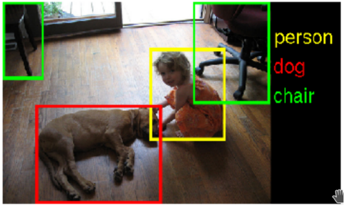
- Captioning:
  "In a very nice spring morning, people having a walk in the sunny beach of Cefalu"

Recognition and classification

- Classification: data → set of pre-defined classes
- Recognition much more general than classification, *e.g.*
  - Object Localization in images
  - Sequence prediction for text, speech, audio, *etc*
- Many tasks can be cast as classification problems

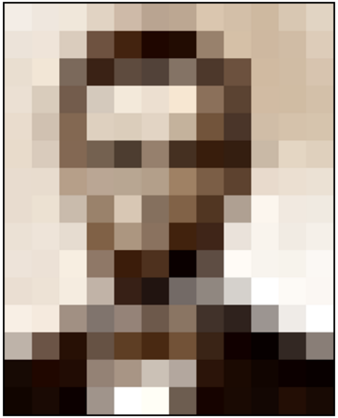
⇒ **Importance of classification**





Recognition of low-level signals: filling the semantic gap

- What we perceive *vs*
  What a computer sees

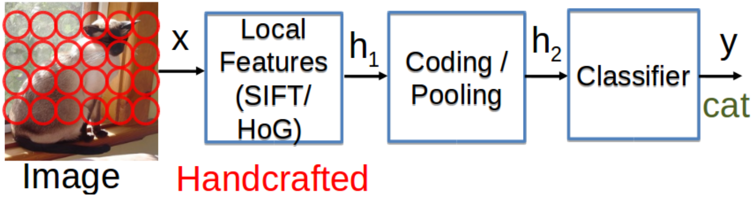


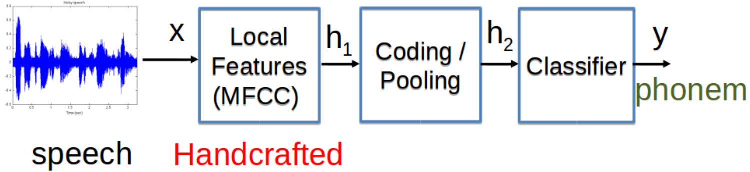
243	239	240	225	206	185	188	218	211	206	216	225
242	239	218	110	67	31	34	152	213	206	208	221
243	242	123	58	94	82	132	77	108	208	208	215
235	217	115	212	243	236	247	139	91	209	208	211
233	208	131	222	219	226	196	114	74	208	213	214
232	217	131	116	77	150	69	56	52	201	228	223
232	232	182	186	184	179	159	123	93	232	235	235
232	236	201	154	216	133	129	81	175	252	241	240
235	238	230	128	172	138	65	63	234	249	241	245
237	236	247	143	59	78	10	94	255	248	247	251
234	237	245	193	55	33	115	144	213	255	253	251
248	245	161	128	149	109	138	65	47	156	239	255
190	107	39	102	94	73	114	58	17	7	51	137
23	32	33	148	168	203	179	43	27	17	12	8
17	26	12	160	255	255	109	22	26	19	35	24

Recognition of low-level signals: input data variations

- Illumination variations
- View-point variations
- Deformable objects
- intra-class variance

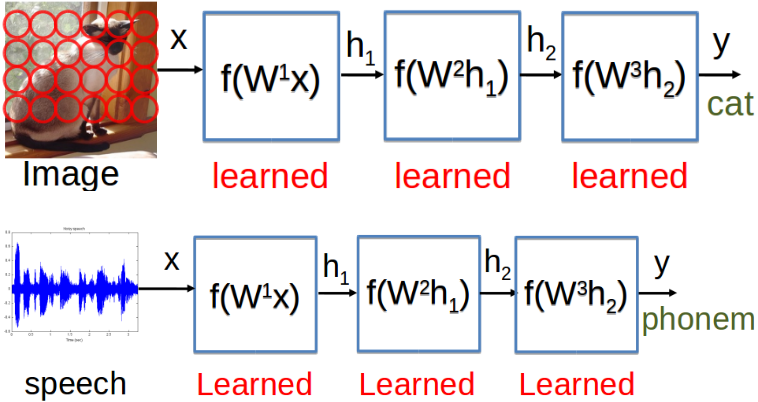
Deep Learning (DL) & Recognition of low-level signals





- Before DL: **handcrafted intermediate representations**
  - ⊖ Needs expertise in each field
  - ⊖ **Shallow archis**: low-level features

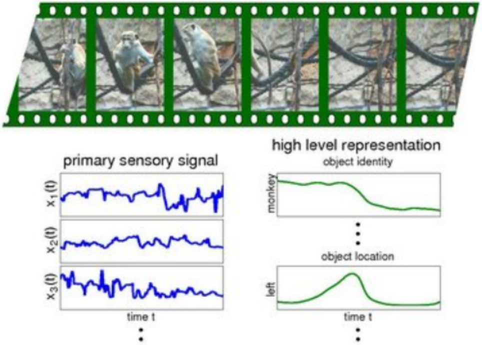
Deep Learning (DL) & Recognition of low-level signals



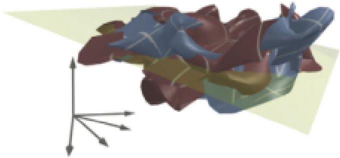
- **DL: learning intermediate representations**
  - ⊕ **Deep**: hierarchy, gradual learning
  - ⊕ Common learning methodology, no expertise

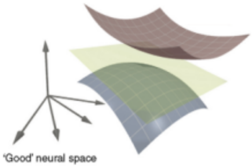
Perception *vs* Acquisition

- Measurements - primary sensory signal: noisy
- Perception - high level representation, *i.e.* object class: stable



Deep Learning (DL) & Manifold Untangling





Raw data:
very tangled manifold

Deep Learning representations:
untangled manifold

- Manifold untangling: neuroscience terminology
- Deep Learning models gradually disentangle data manifold
- Deformations linearized: simple linear classifier
  in disentangled DL manifold space!

Deep Learning Context: Conclusion

- Deep Learning: Representation Learning *vs* feature engineering
- Manifold untangling: stable representations
- **Computational model for neural networks?**
  $\Rightarrow$ following!