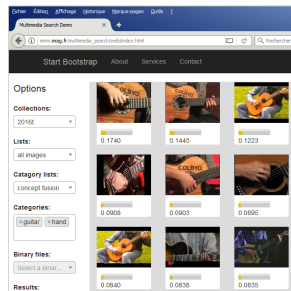


## Deep Convolutional and Recurrent networks for image, speech & text

- ▶ Success of deep learning in vision, speech, NLP, ...
  - ▶ Many processing layers from raw input signal upwards
  - ▶ All parameters trained jointly in end-to-end manner
  - ▶ Why now: Data, computation, training algorithms
- ▶ Areas of research in DeCore
  - ▶ Multi-modal data: modeling relation images and text
  - ▶ Visual recognition: many classes, incremental learning
  - ▶ Time series with multiple resolutions and missing data
  - ▶ Automating network architecture design
  - ▶ Processing non-regular data: 3D shapes, molecular graphs, etc.
- ▶ Who?
  - ▶ Organizers: L. Besacier, D. Pellerin, G. Quénot, J. Verbeek
  - ▶ Labs: GIPSA, LIG, LJK
  - ▶ Teams: AGPIG, AMA, GETALP, MRIM, SigmaPhy, THOTH

# PhD theses funded by DeCoRe

- ▶ Anuvabh Dutt: Object recognition and localization
  - ▶ Supervision: D. Pellerin & G. Quénot
  - ▶ Hierarchical concept recognition
  - ▶ Network architecture evolution for incremental learning
  - ▶ Concept-based multimedia retrieval
- ▶ Maha Elbayad: Natural language image description
  - ▶ Supervision: L. Besacier & J. Verbeek
  - ▶ CNN-RNN design from pixels to words
  - ▶ Increase diversity in generated captions
  - ▶ Visual attention for compositionality



Search in videos: "Guitar and Hand"



Output: "A young boy is holding a cell phone."