AI in Medicine

Working Group Discussion Session, APMSS 2019

Chair: 謝德威 Alexander Te-Wei Shieh

Agenda

Introduction

1. Machine Learning and Information Retrieval
2. Computer Vision / Natural Language Processing
3. Reinforcement Learning and Unsupervised Learning
4. Current AI applications in medicine
5. Some Challenges in medical AI

Discussion

(0) Self-Introduction

謝德威 Alexander Te-Wei Shieh

NTU Medicine, 3rd year

NTU Computer Science and Information Engineering (double major)

My interests are: Natural Language Processing (and related ML techniques), Tech Startups.

1. Cheminformatics / Drug Development (Computational Molecular Design and Metabolomics Lab)
2. Natural Language Processing (Machine Intelligence and Understanding Lab)

(0) Self-Introduction

1. Name
2. School / n-th year
3. Why you choose this topic?
4. (Any previous experience related to AI)

Memoji by Apple

(0) AI in Medicine

This session will be extremely fast-paced.
If you have any questions please shout out!

Today our discussion is much more limited:

Artificial General Intelligence (Strong AI)

Artificial Narrow Intelligence **(Weak AI)**

- **Machine Learning**
  - **Deep Learning**
  - Probabilistic Models, SVMs, Random Forests, ...
- Robotics, Search, Genetic Algorithms, ...

(1) Machine Learning and Information Retrieval

Information Retrieval: Vector-Space Model

How did PubMed find your article?

Indexing, Retrieval Model, Query-Expansion, Re-Rank (Relevance Feedback)

Document Vector

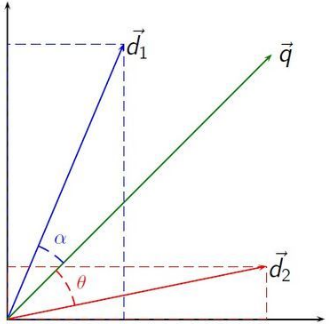
$$\mathbf{v}_d = [w_{1,d}, w_{2,d}, \ldots, w_{N,d}]^T$$

$$w_{t,d} = \mathrm{tf}_{t,d} \cdot \log \frac{|D|}{|\{d' \in D \,|\, t \in d'\}|}$$

TF-IDF Weighting

Cosine Similarity
between Document Vector and Query Vector

$$\cos(d_j, q) = \frac{\mathbf{d_j} \cdot \mathbf{q}}{\|\mathbf{d_j}\| \, \|\mathbf{q}\|} = \frac{\sum_{i=1}^{N} w_{i,j} w_{i,q}}{\sqrt{\sum_{i=1}^{N} w_{i,j}^2} \sqrt{\sum_{i=1}^{N} w_{i,q}^2}}$$



TREC precision medicine / clinical support track: http://www.trec-cds.org/2018.html

(1) Machine Learning and Information Retrieval

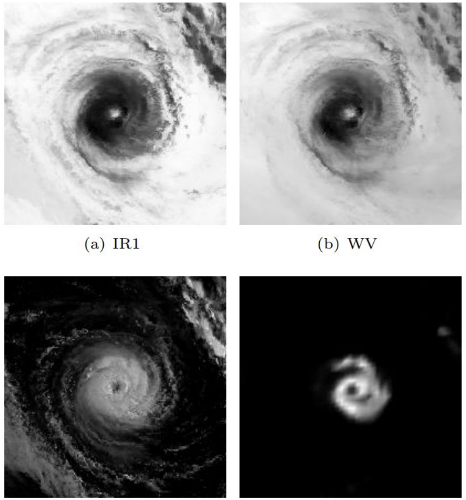
Why machine learning?

Machine Learning Problems

1. **Supervised** Learning [Labeled]
    a. Regression: [Compound, Activity]
    b. Classification: [Image, Benign / Malignant], [Text, ICD Codes]
2. **Unsupervised** Learning [Unlabeled]
    a. Representation Learning: Word2Vec, Autoencoders, Language Models
    b. Clustering
3. **Reinforcement** Learning [Reward]

(1) Machine Learning and Information Retrieval

Supervised Learning: Regression (Convolutional Neural Networks)



(a) IR1 (b) WV

(c) VIS (d) PMW

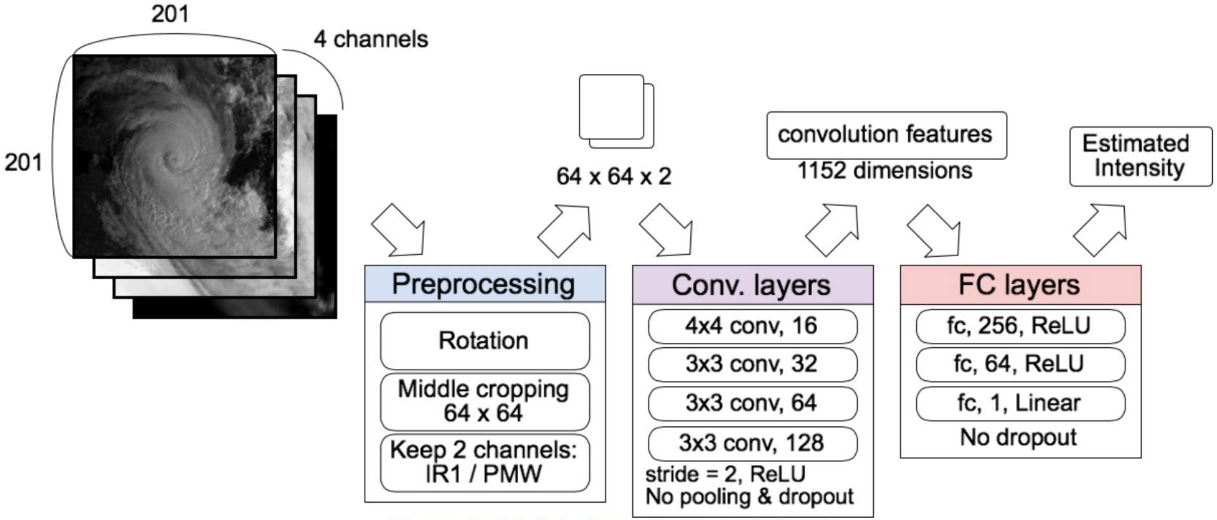


Figure 2: Model structure for CNN-TC

Chen et al. **Rotation-blended CNNs** on a New Open Dataset for **Tropical Cyclone Image-to-intensity Regression**. KDD 2018.

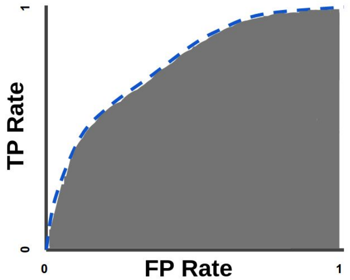
(1) Machine Learning and Information Retrieval

Evaluation for Regression Problems

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \quad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

- Root Mean Squared Error (**RMSE**), Mean Average Error (**MAE**)

Evaluation for Classification Problems



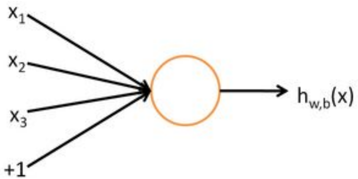
- **Accuracy** (*(TP+TN)/(P+N)*)
- **Sensitivity** (=Recall/**TPR**, *TP/P*) / **Specificity** (=**TNR**, *TN/N*)
- **Precision** (=**PPV**, *TP/(TP+FP)*) / **Recall** / **F1** (*2PPV×TPR/(PPV+TPR)*)
- **ROC curve** (TPR vs TNR at different thresholds) **/** Area Under ROC curve (**AUC**)

Others: MAP, BLEU & ROUGE, Perpelexity, Human-evaluation, Task-specific scores

(1) Machine Learning and Information Retrieval

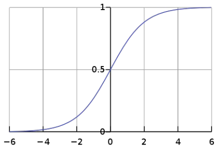
From Logistic Regression to **Deep Learning**

Logistic Regression

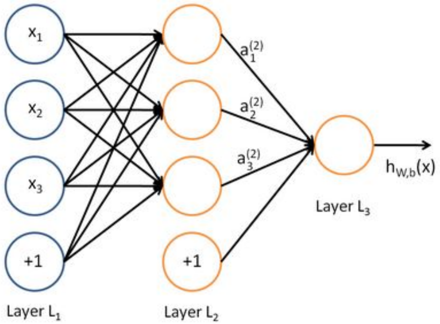


$$h_{W,b}(x) = f(W^T x) = f(\sum_{i=1}^{3} W_i x_i + b)$$

$$f(z) = \frac{1}{1+\exp(-z)}.$$



Multi-Layer Perceptron



$$\begin{aligned}
a_1^{(2)} &= f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)}) \\
a_2^{(2)} &= f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)}) \\
a_3^{(2)} &= f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)}) \\
h_{W,b}(x) = a_1^{(3)} &= f(W_{11}^{(2)}a_1^{(2)} + W_{12}^{(2)}a_2^{(2)} + W_{13}^{(2)}a_3^{(2)} + b_1^{(2)})
\end{aligned}$$

$$\begin{aligned}
z^{(2)} &= W^{(1)}x + b^{(1)} \\
a^{(2)} &= f(z^{(2)}) \\
z^{(3)} &= W^{(2)}a^{(2)} + b^{(2)} \\
h_{W,b}(x) = a^{(3)} &= f(z^{(3)})
\end{aligned}$$

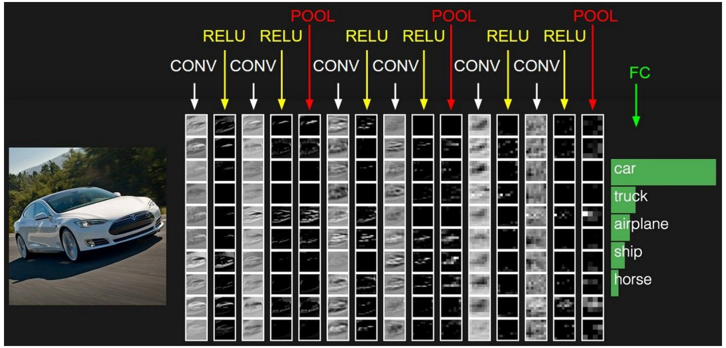
Matrix Form

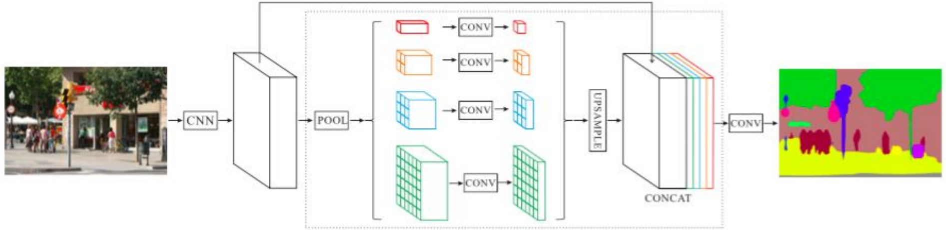
From: http://ufldl.stanford.edu/wiki/index.php/Neural_Networks

(2) CV and NLP A deep learning point of view

Convolutional Neural Networks (CNN)

- http://cs231n.github.io/convolutional-networks/
- http://blog.qure.ai/notes/semantic-segmentation-deep-learning-review

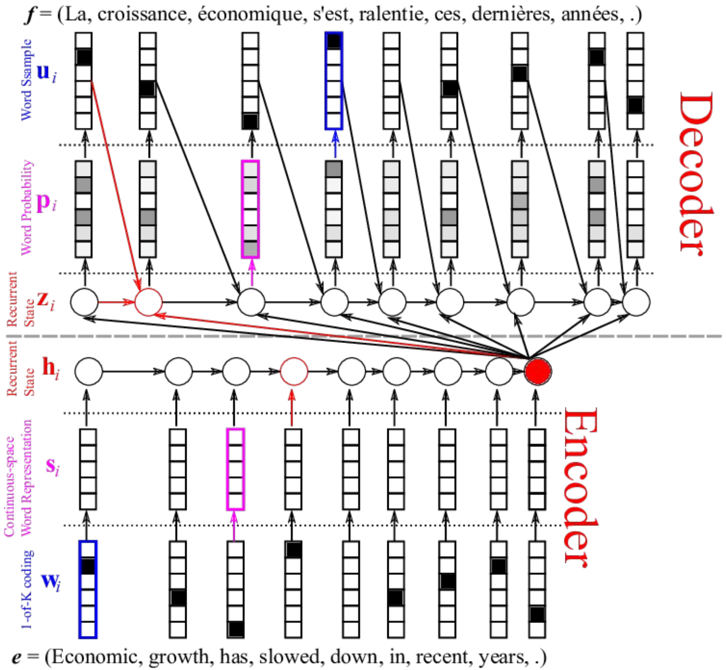




(a) Input Image (b) Feature Map (c) Pyramid Pooling Module (d) Final Prediction

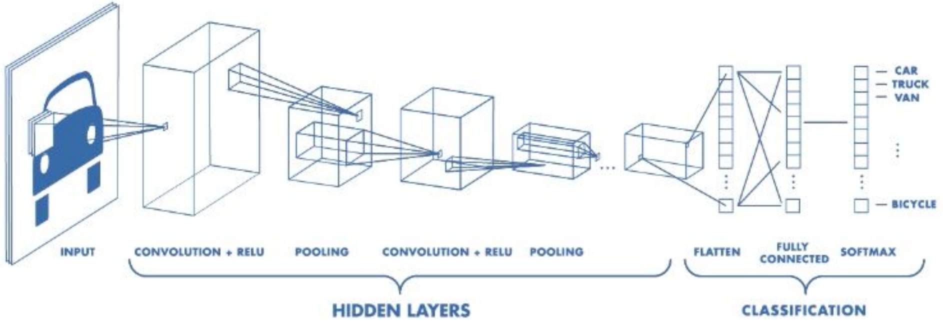
Seq2Seq with Recurrent Neural Networks (RNN)

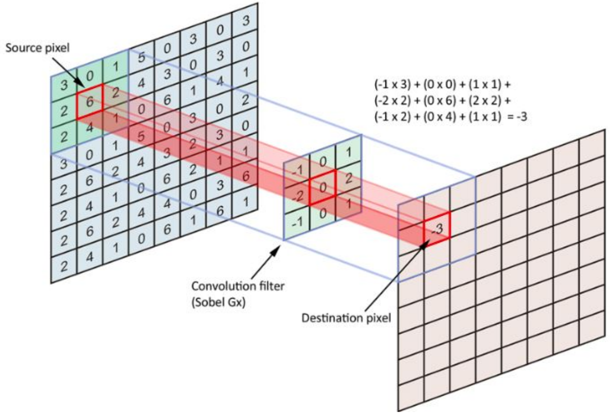
https://devblogs.nvidia.com/introduction-neural-machine-translation-gpus-part-2/

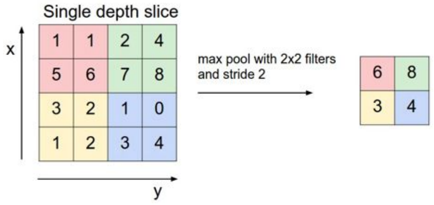


(2) CV and NLP A deep learning point of view

Convolutional Neural Network (CNN)



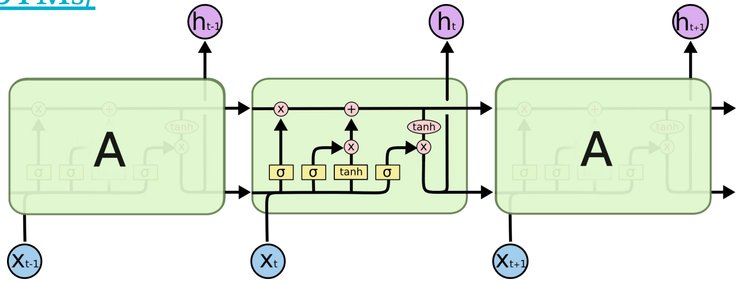


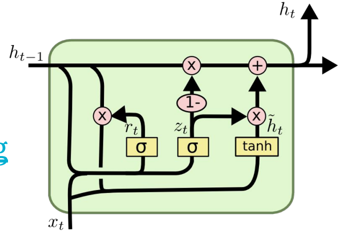


https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050

Recurrent Neural Network (RNN) with **Long Short Term Memory (LSTM)** Units or **Gated Recurrent Units (GRU)**

http://colah.github.io/posts/2015-08-Understanding-LSTMs/





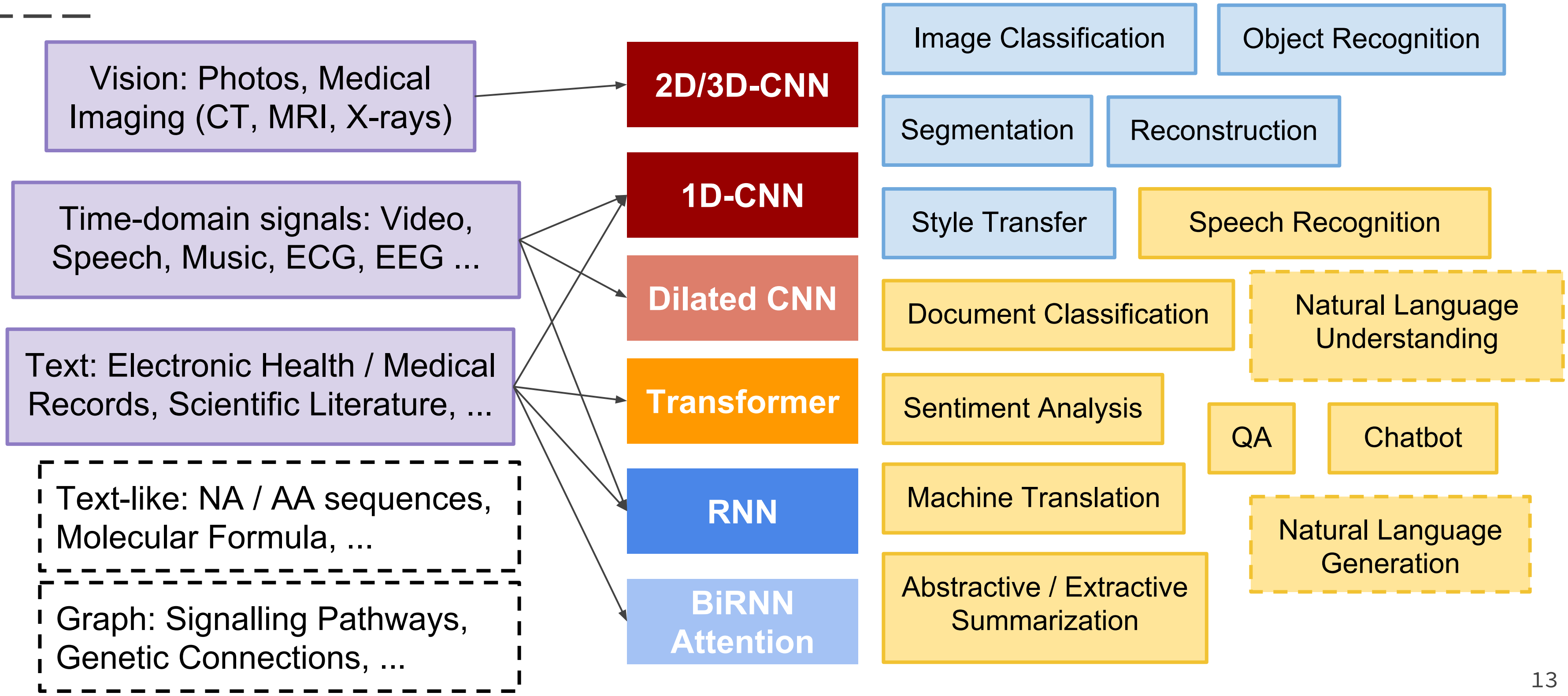
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

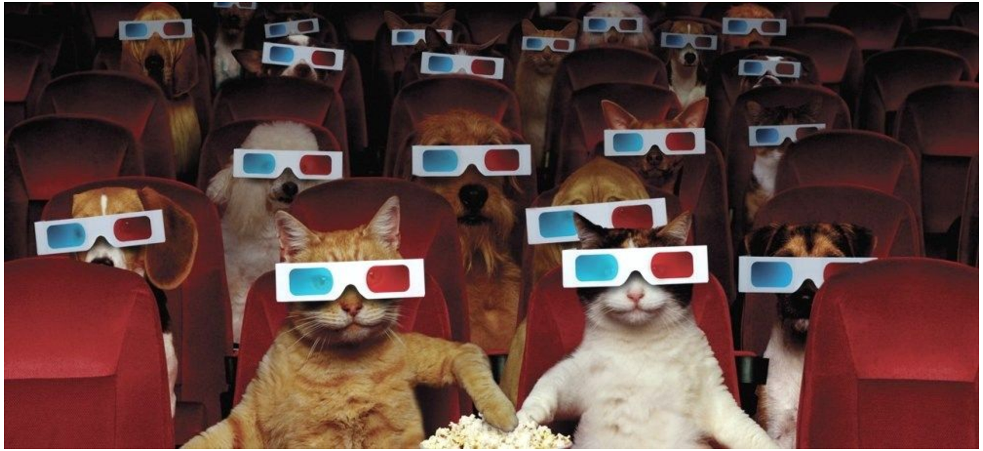
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

(2) CV and NLP A deep learning point of view



(2) CV and NLP

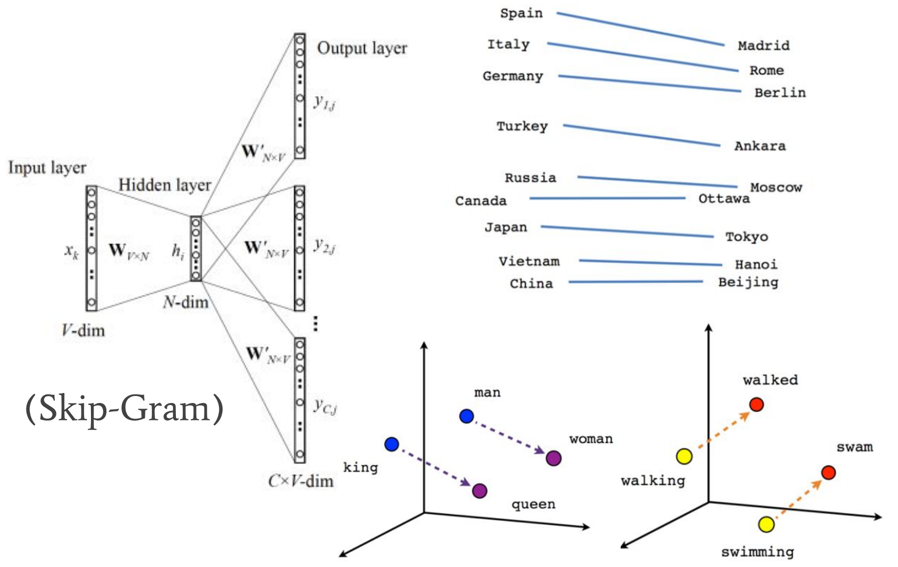
Demos



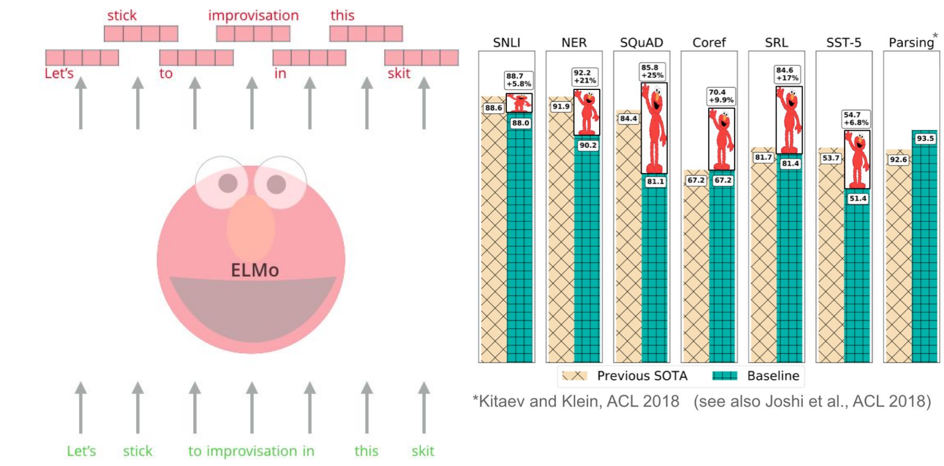
- Multimodal Unsupervised Image-to-image Translation, https://youtu.be/ab64TWzWn40
- A Style-Based Generator Architecture for Generative Adversarial Networks (GAN), https://youtu.be/kSLJriaOumA
- Image Inpainting for Irregular Holes Using Partial Convolutions, https://youtu.be/gg0F5JjKmhA
- Google's AI Assistant Can Now Make Real Phone Calls (Task-Oriented Chatbot, Speech Recognition, Natural Language Understanding, Text-to-Speech), https://youtu.be/JvbHu_bVa_g

(3) RL and Unsupervised

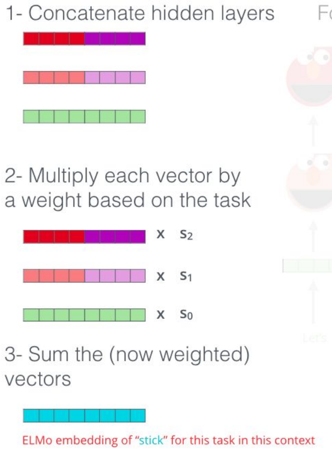
Unsupervised Learning: Word2Vec vs Elmo

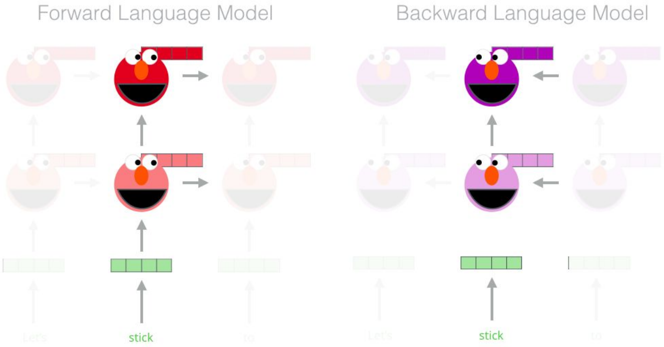


(Skip-Gram)



*Kitaev and Klein, ACL 2018 (see also Joshi et al., ACL 2018)



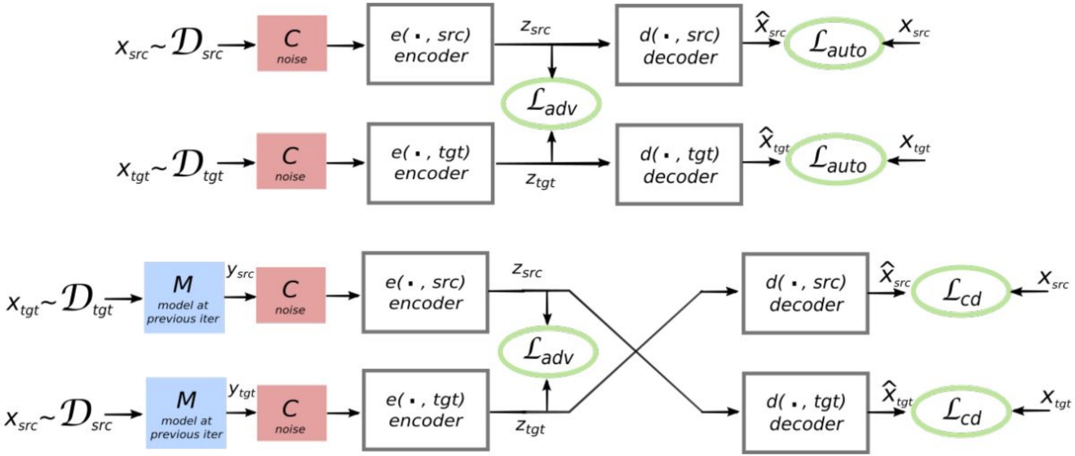


- http://jalammar.github.io/illustrated-bert/
- http://ruder.io/nlp-imagenet/

(3) RL and Unsupervised

Unsupervised Learning: Generative Adverserial Networks

- Huang et al. [Multimodal Unsupervised Image-to-Image Translation](). ECCV 2018.
- Lample et al. [Unsupervised Machine Translation Using Monolingual Corpora Only](). ICLR 2018.



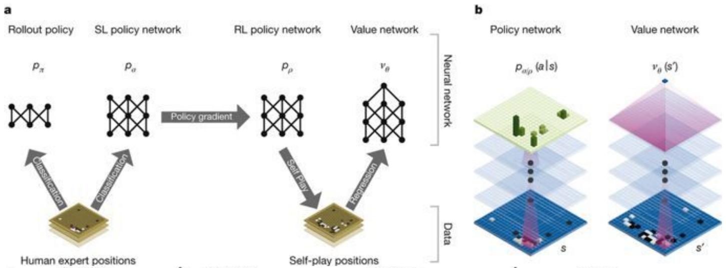
(a) edges → shoes
(b) big cats → house cats

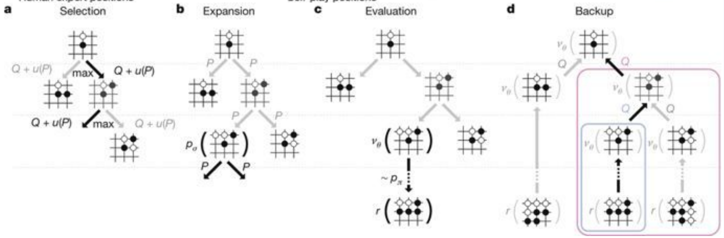
(3) RL and Unsupervised

Reinforcement Learning: Atari DQN and AlphaGo

AlphaGo

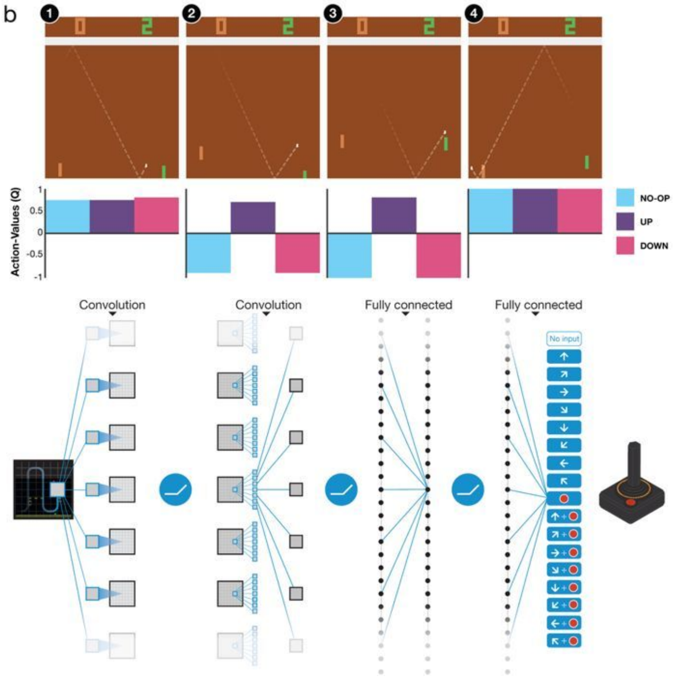
Policy / Value Network
Monte-Carlo Tree Search





Atari DQN

Deep Q-Learning (From CNN to Action)
Master 49 Games



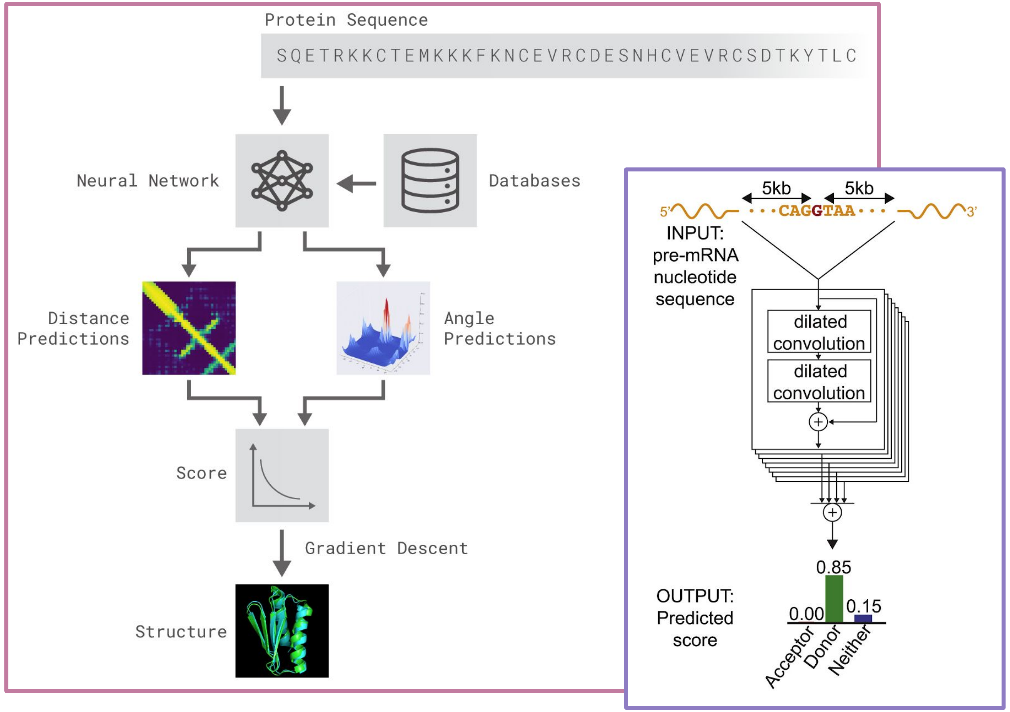
- Mnih et al. Human-level control through deep reinforcement learning. Nature 2015.
- Silver et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016.
- https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/

(4) Current AI Applications in Medicine

Structural / Molecular Biology

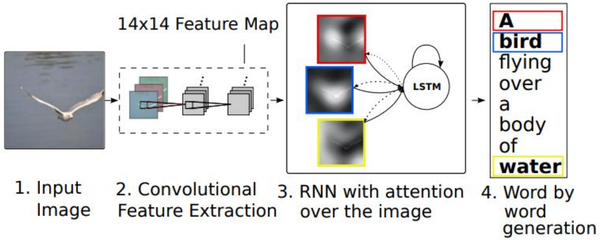
Predict **Protein Folding**: AlphaFold (DeepMind).

Predict **Pre-mRNA Splicing**: Jaganathan et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell 2019.

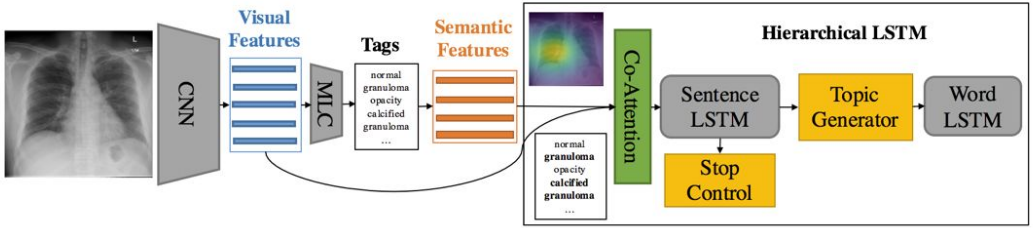


(4) Current AI Applications in Medicine

Xu et al. [Show, Attend and Tell: Neural Image Caption Generation with Visual Attention](). ICML 2015.



Jing et al. [On the Automatic Generation of Medical Imaging Reports](). ACL 2018.

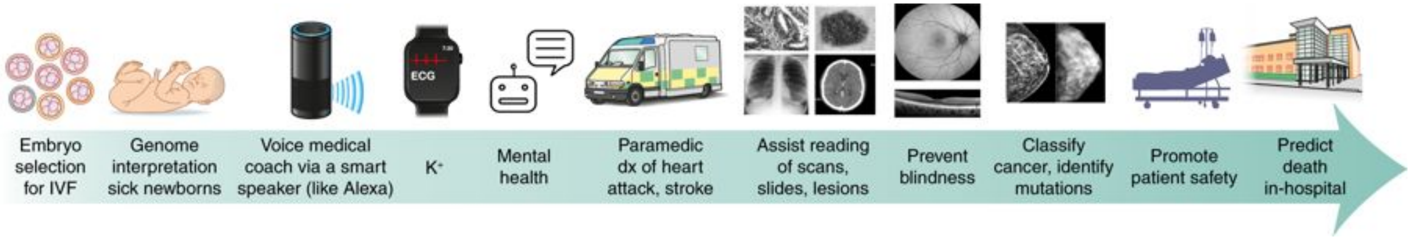


Ground Truth	Ours-CoAttention
No active disease. The heart and lungs have in the interval. Both lungs are clear and expanded. Heart and mediastinum normal.	No active disease. The heart and lungs have in the interval. Lungs are clear and expanded. Cardiomediastinal silhouette is within normal limits. No pleural effusion or pneumothorax is seen. No pleural effusion. No cavitary or pneumothorax.
No evidence of active disease. The lungs are clear. There is no focal airspace consolidation. No pleural effusion or pneumothorax. Heart size and mediastinal contour are within normal limits. There are multilevel degenerative changes of the spine.	No acute cardiopulmonary findings. Heart size is not enlarged. No focal airspace consolidation suspicious pulmonary opacity large pleural effusion or pneumothorax. No focal areas of consolidation. Degenerative changes of the spine. This is moderate exam of the hydropneumothorax. Lungs are clear. There is no focal airspace consolidation pleural effusion or pneumothorax.
No acute cardiopulmonary abnormality. Normal heart size mediastinal contours. Eventration of the right hemidiaphragm. No focal airspace consolidation. No pleural effusion or pneumothorax.	No acute cardiopulmonary abnormality. Stable appearance of the thoracic aorta. The right lateral lower lobe is noted in the right lower right midlung. No large pleural effusion or focal airspace disease. Mild interstitial opacities. Atherosclerotic calcifications bony structures bilaterally. There is no pleural effusion or pneumothorax developed in the right lower lobe.
No acute cardiopulmonary abnormality. Heart size appears within normal limits . Pulmonary vasculature appears within normal limits. Overlying the middle cardiac silhouette representing a hiatal hernia. No focal consolidation pleural effusion or pneumothorax. No acute bony abnormality.	No active disease. The heart and lungs have in the interval. Nipple and lateral lucency in the lungs suggestive of focal airspace disease. The lungs are hyperexpanded consistent with emphysema in the left lower lobe. This is most at the upper lobes. This may indicate hypoventilated irregularities or effusions. The lungs are otherwise grossly clear. Resolution of by normal pleural effusion.

(4) Current AI Applications in Medicine

Medical Diagnosis, Prognosis

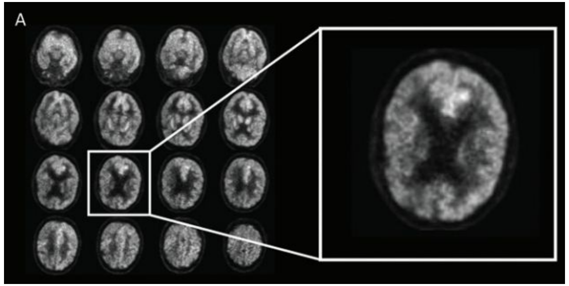
Demo Paper: Ding et al. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. Radiology 2018.



Recommended Reading: Topol. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine 2019.

(4) Current AI Applications in Medicine

- Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD)
- Alzheimer’s Disease Neuroimaging Initiative (ADNI) (2109 imaging studies / 1002 patients) and independent test set (40 imaging studies / 40 patients), Size: 512×512
- Model: Inception V3 CNN



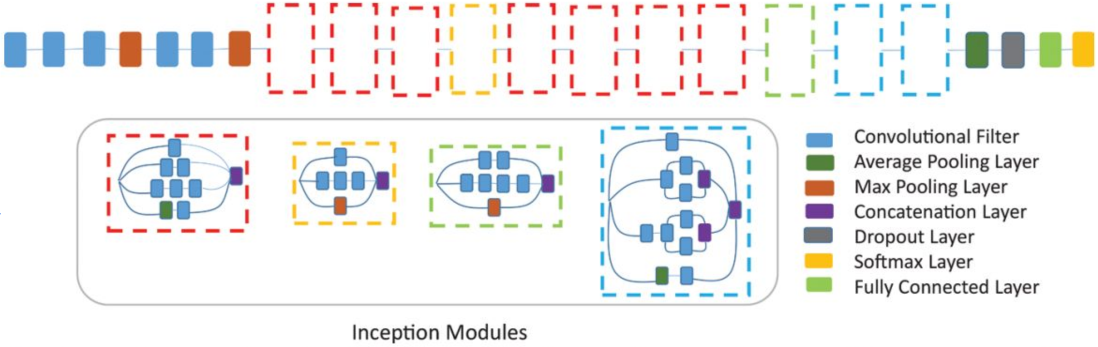


Figure 3: Convolutional neural network architecture, Inception v3, used in this study. Inception v3 network stacks 11 inception modules where each module consists of pooling layers and convolutional filters with rectified linear units as activation function. The input of the model is two-dimensional images of 16 horizontal sections of the brain placed on 4 × 4 grids as produced by the preprocessing step. Three fully connected layers of size 1024, 512, and 3 are added to the final concatenation layer. A dropout with rate of 0.6 is applied before the fully connected layers as means of regularization. The model is pretrained on ImageNet dataset and further fine-tuned with a batch size of 8 and learning rate of 0.0001.

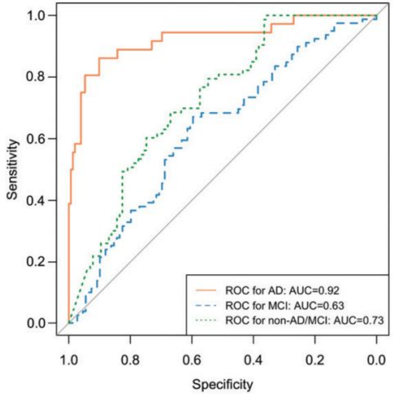
(4) Current AI Applications in Medicine

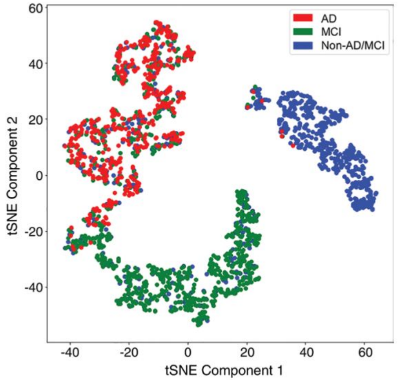
- AUC = 0.98 [0.94, 1.00], specificity = 82% when sensitivity = 100%
- (average of 75.8 months prior to the final diagnosis.)

Table 2: Performance Comparison of Deep Learning Algorithm and Radiology Readers

Parameter	Sensitivity (%)*	Specificity (%)*	Precision (%)*	F1 Score (%)	No. of Imaging Studies
Deep learning model on 10% ADNI set					
AD	81 (29/36)	94 (143/152)	76 (29/38)	78	36
MCI	54 (43/79)	68 (74/109)	55 (43/78)	55	79
Non-AD/MCI	59 (43/73)	75 (86/115)	60 (43/72)	59	73
Deep learning model on independent test set					
AD	100 (7/7)†	82 (27/33)	54 (7/13)	70†	7
MCI	43 (3/7)†	58 (19/33)	18 (3/17)†	25†	7
Non-AD/MCI	35 (9/26)	93 (13/14)†	90 (9/10)†	50	26
Radiology readers on independent test set					
AD	57 (4/7)	91 (30/33)	57 (4/7)	57	7
MCI	14 (1/7)	76 (25/33)	11 (1/9)	13	7
Non-AD/MCI	77 (20/26)	71 (10/14)	83 (20/24)	80	26

Note.–Unless otherwise indicated, data are averages ± standard deviation. ADNI = Alzheimer's Disease Neuroimaging Initiative, AD = Alzheimer disease, MCI = mild cognitive impairment, Non-AD/MCI = neither Alzheimer disease nor mild cognitive impairment.
* Numbers in parentheses are raw data used to calculate the percentage.
† Numbers indicate higher performance from deep learning algorithm compared with reader performance on independent test set.





Key Takeaways

1. CNNs and RNNs are prevalent methods in deep learning
2. Many medical AI applications borrowed ideas for CV and NLP.
3. Use the data to find appropriate algorithms! (but not vice versa)
4. Solving a problem do not always require advanced AI / deep learning or even machine learning.
5. Use domain knowledge to establish good inductive bias.
6. Will doctors be replaced by AI? Try to build an AI and lets see!

Learn More

Machine Learning DIY: Scikit-Learn, Tensorflow(Keras), PyTorch, Google Cloud ML

Google: **Machine Learning Crash Course** https://developers.google.com/machine-learning/crash-course/

MLDS lectures by Prof. Hung-Yi Lee @ NTU EE (in Chinese)
http://speech.ee.ntu.edu.tw/~tlkagk/courses_MLDS18.html

Machine Learning for Health Workshop, NIPS 2018 https://ml4health.github.io/2018/pages/papers.html

AI for Social Good Workshop, NIPS 2018 https://aiforsocialgood.github.io/2018/

Stanford ML Group https://stanfordmlgroup.github.io/

Follow some works by Google Brain, DeepMind, Facebook AI Research, NVIDIA Research, ...

Discussion

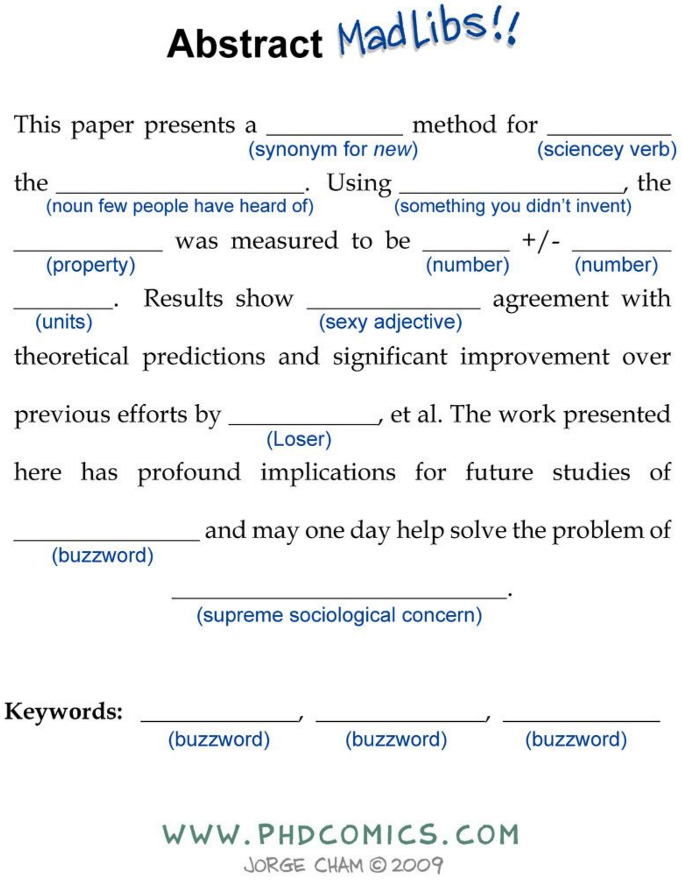
---

1. Lecouat et al. Semi-Supervised Deep Learning for Abnormality Classification in Retinal Images. ML4H Workshop, NIPS 2018.
2. Hannun et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine 2019. project site
3. Ribeiro et al. Automatic Diagnosis of Short-Duration 12-Lead ECG using a Deep Convolutional Network. ML4H Workshop, NIPS 2018.
4. Komorowski et al. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nature Medicine 2018. related work 1, related work 2

Discussion

<u>**Paper Presentation: 3 min, be concise !!!**</u>

1. What's the **objective** of this AI application?
2. Describe the **dataset**, including size, dimensions, endpoints, etc.
3. Describe the authors' **approach**: which algorithm did they use and why?
4. Describe the **evaluation** method and results.
5. (Opt) What's the significance of this work?
6. (Opt) What are the limitations discussed by the authors?



Discussion

Design a medical AI application

1. What's the **objective** of this AI application?
2. Where is the **data**?
3. How to collect the desired data?
4. What kind of **model** will you choose and why?
5. What is the **evaluation criteria** for your model?

Thank you!
Any Questions?

If you have further inquiries, please email me:
b05401009@ntu.edu.tw