Lecture 10: Multimodal data, multimodal models, and weakly supervised learning

Announcements

- HW 1 and project proposal grades have been released
- Upcoming deadlines:
  - A2 due today Wed Oct 21
- Project milestone due Fri Oct 30
- Project milestone presentations Mon Nov 2 in-class
  - 4 minutes per group, strict time limit. It's ok to have a subset of group members present
  - Should summarize all components of milestone report (5 pts total)
  - Pre-recorded video option can be requested for those unable to attend
  - See Piazza post for more details about all of this

Today

- One more example of deep learning in genomics
- Multimodal data and models
- Weakly supervised learning

Remember: ChIP-seq

Produces reads of DNA sequences where a protein binds

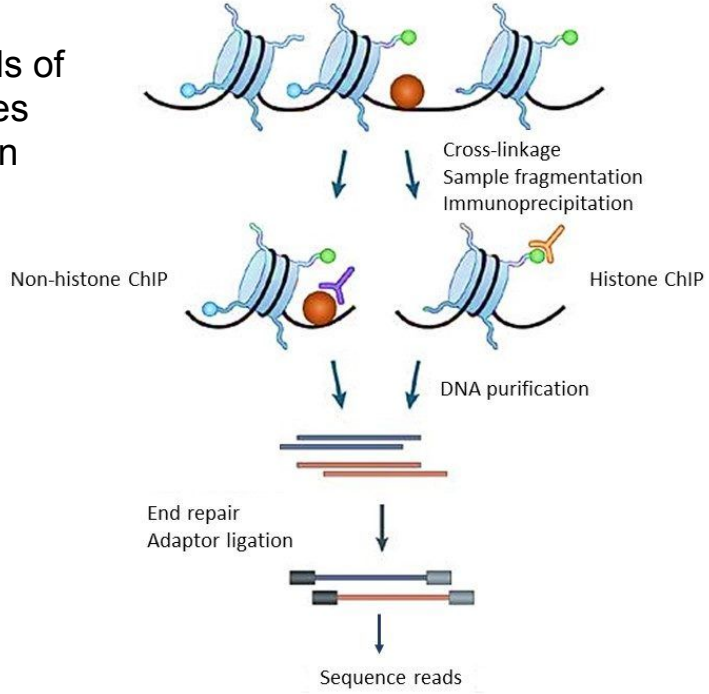


Figure credit:
https://www.france-genomique.org/wp-content/uploads/2019/08/CHIP-selon-Park-1-e1566900408602.jpg

Visualize distribution of locations on DNA where protein binds

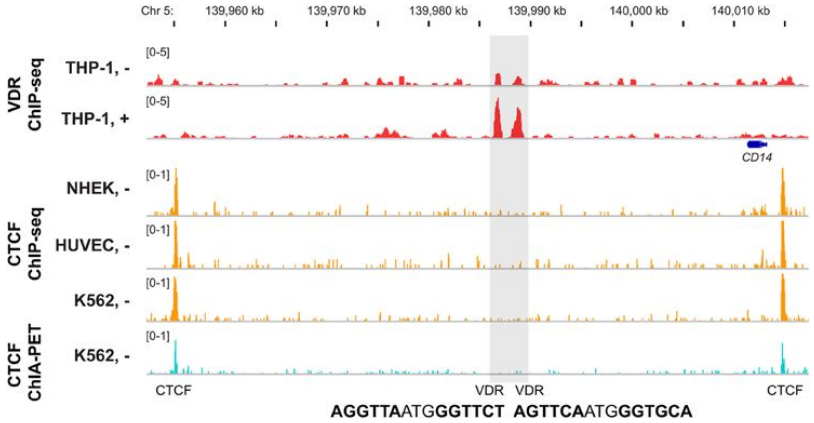
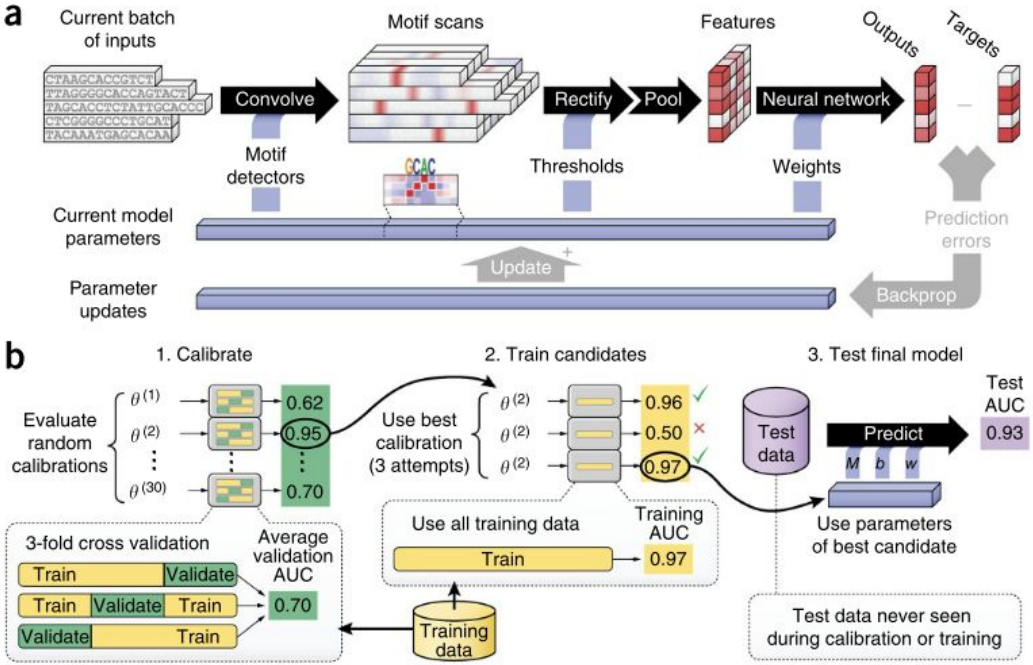


Figure credit:
https://www.researchgate.net/publication/262150050/figure/fig2/AS:272566950559751@1441996433141/Chromatin-domain-containing-VDR-binding-sites-The-IGV-browser-was-used-to-display-the.png

Remember: DeepBind

Input: DNA sequence
Output: Score of whether a particular protein will bind to the sequence or not

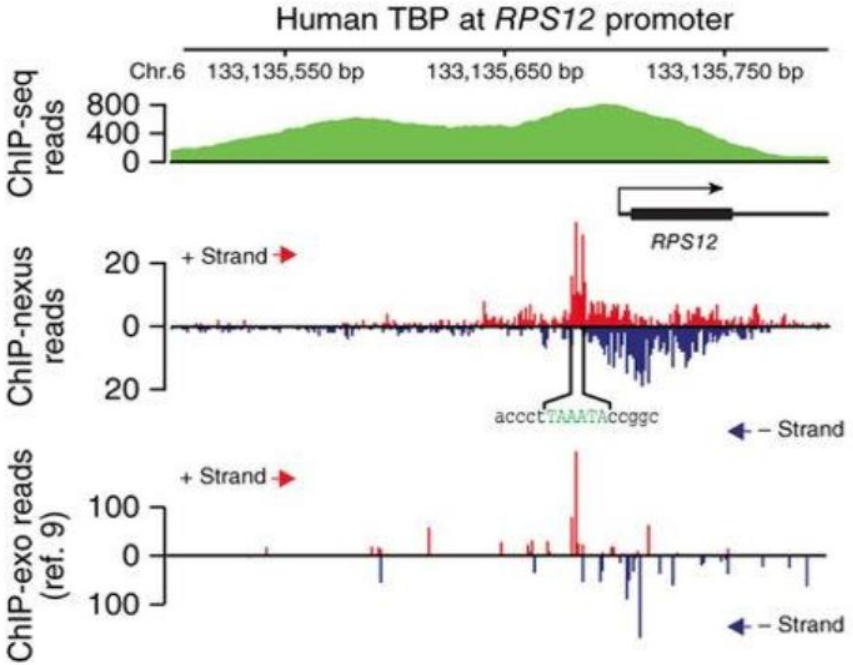
- Processing to handle different sources of experimental (training) data and input / output data formats
- Trained on 12 TB of sequence data; learned 927 DeepBind models representing 538 transcription factor (TF) proteins and 194 RNA-binding proteins (RBPs)



Alipanahi et al. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature Biotechnology, 2015.

More recently: ChIP-nexus vs. ChIP-seq

ChIP-nexus: newer technology that enables improved and higher-resolution data about transcription factor binding footprints on DNA (at individual base-pair resolution)

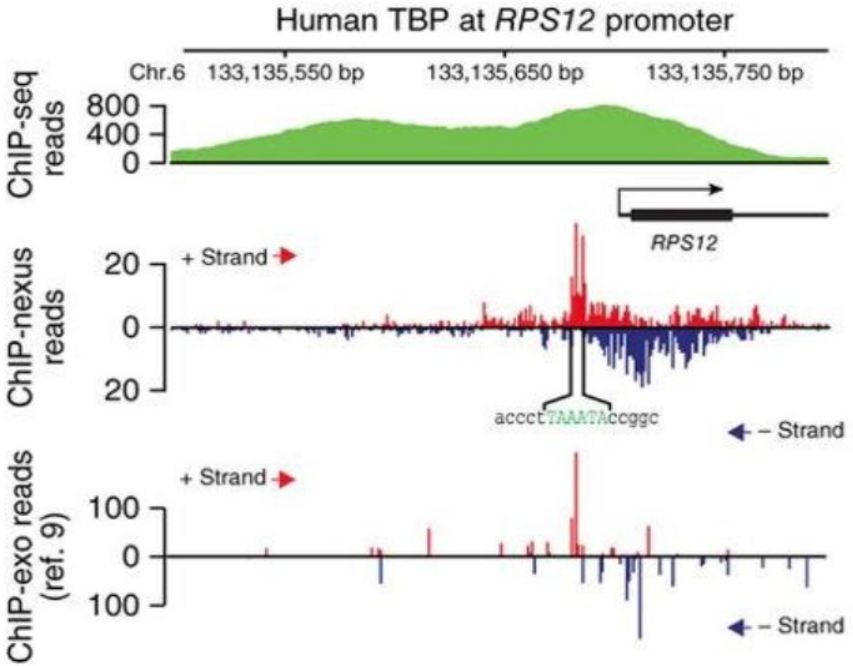


He et al. ChIP-nexus enables improved detection of *in vivo* transcription factor binding footprints. Nature Biotechnology, 2015.

More recently: ChIP-nexus vs. ChIP-seq

ChIP-nexus: newer technology that enables improved and higher-resolution data about transcription factor binding footprints on DNA (at individual base-pair resolution)

ChiP-seq

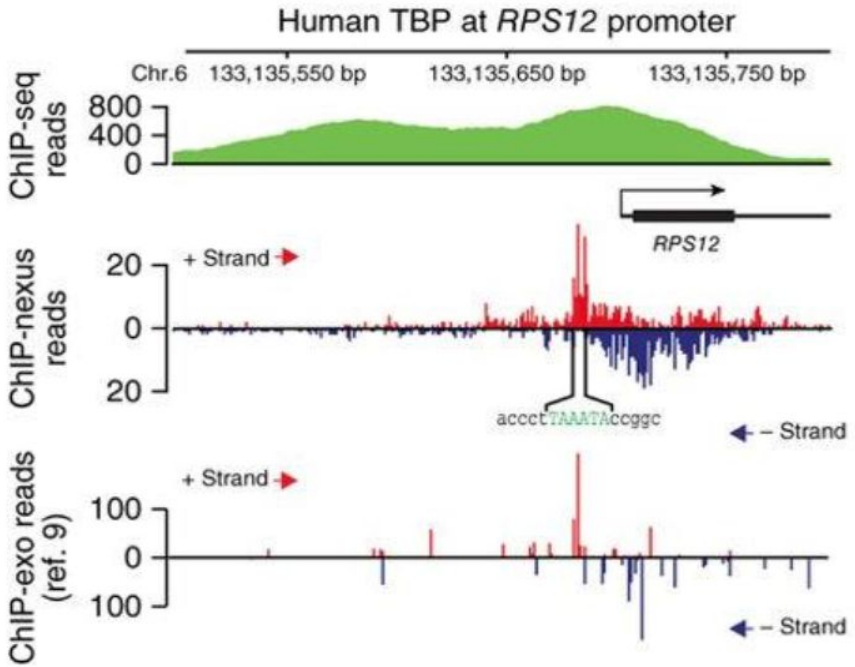


He et al. ChIP-nexus enables improved detection of *in vivo* transcription factor binding footprints. Nature Biotechnology, 2015.

More recently: ChIP-nexus vs. ChIP-seq

ChIP-nexus: newer technology that enables improved and higher-resolution data about transcription factor (TF) binding footprints on DNA (at individual base-pair resolution)

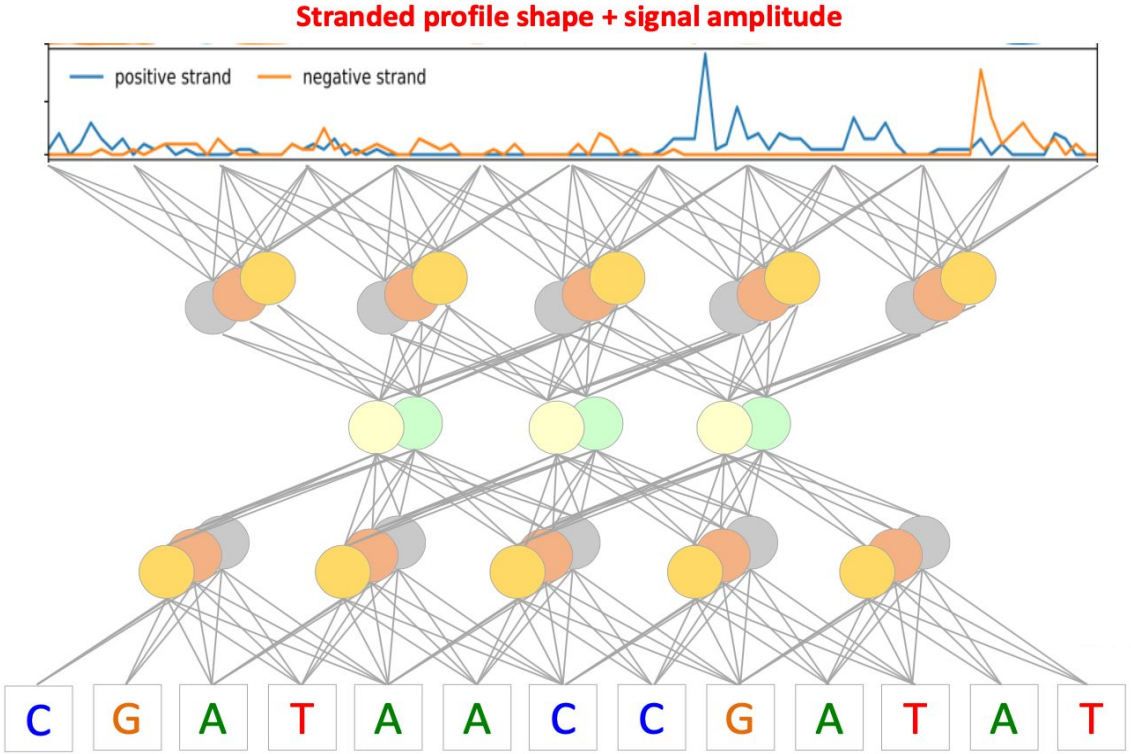
ChiP-nexus



He et al. ChIP-nexus enables improved detection of *in vivo* transcription factor binding footprints. Nature Biotechnology, 2015.

BPNet: DNA sequence to base-pair resolution profile regression

- Deep learning-based model based on ChiP-nexus data, that predicts TF binding profile at high, individual base-pair resolution

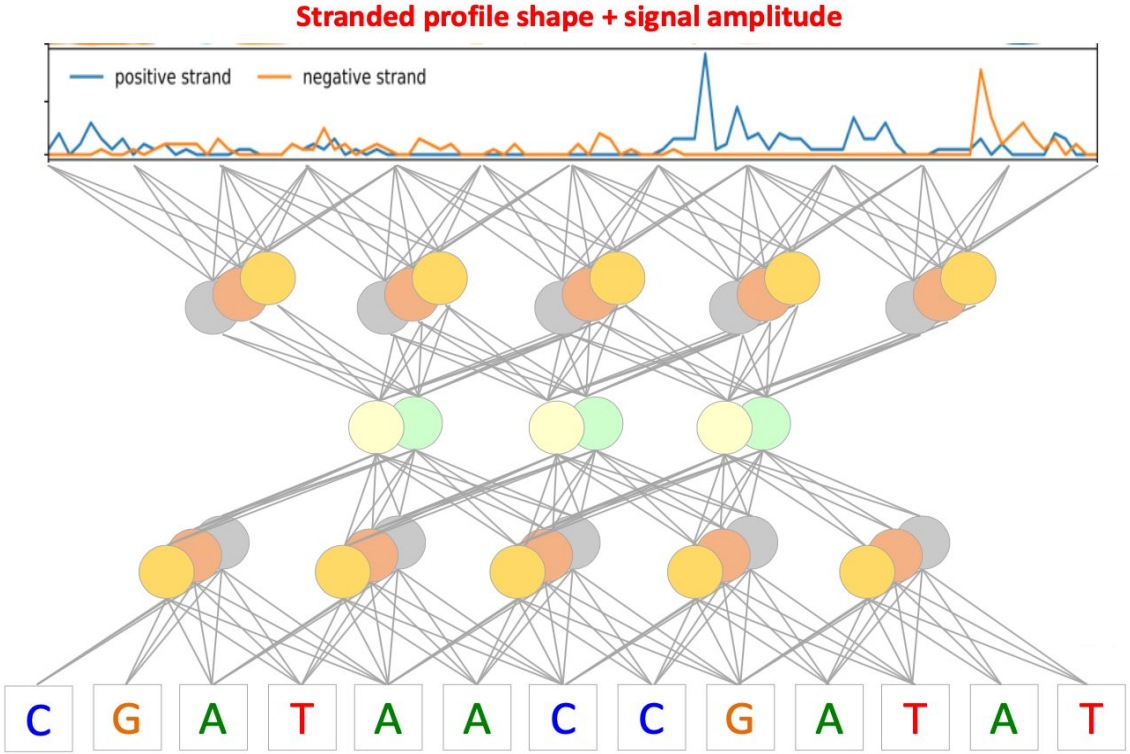


Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.

Slide Credit: Anshul Kundaje

BPNet: DNA sequence to base-pair resolution profile regression

- Deep learning-based model based on ChiP-nexus data, that predicts TF binding profile at high, individual base-pair resolution
- Uses 1-D, **dilated** convolutional layers for greater increase of receptive field (extent of input used to produce a neuron output), instead of pooling layers -> maintains base-pair resolution

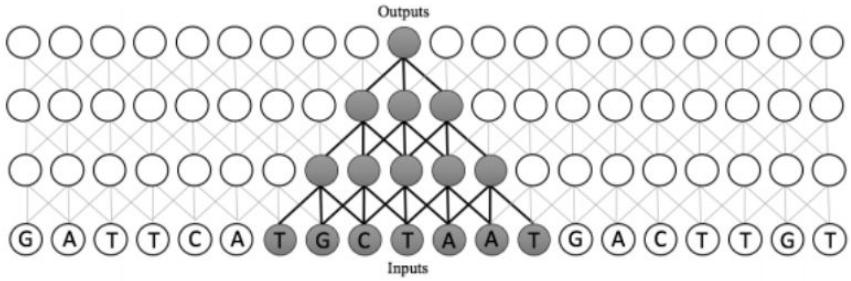


Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.

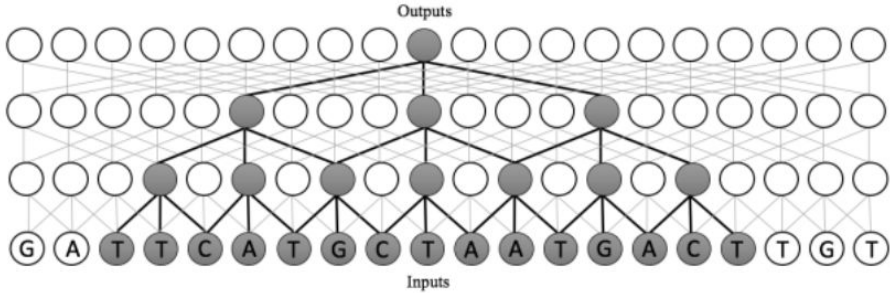
Slide Credit: Anshul Kundaje

Dilated convolutions instead of convolutions

- Greater increase of receptive field vs. standard convolution, for the same # of layers (avoids requiring many layers to increase receptive field which is more difficult to train)
- Pooling layers can also increase receptive field, but reduce resolution (whereas dilated convolutions can maintain high resolution)
- BPNet also includes residual connections (remember ResNets!) to improve ease of optimization for more effective training



(a) Convolution



(c) Dilated Convolution

Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.
Figure credit: Gupta et al. Dilated Convolutions for Modeling Long-Distance Genomic Dependencies, 2017.

Slide Credit: Anshul Kundaje

BPNet: Profile regression loss

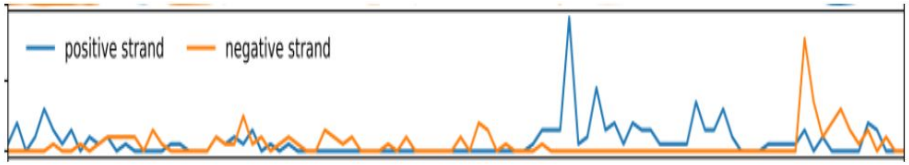
- Two-part loss function for optimizing prediction of the binding profile across the input sequence
  - MSE loss for log (total number of counts across the entire 1kb input sequence)
  - Multinomial loss for the likelihood of the observed count distribution over the sequence, compared to the predicted probabilities

Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.

Slide Credit: Anshul Kundaje

BPNet: Profile regression loss

- Two-part loss function for optimizing prediction of the binding profile across the input sequence
  - MSE loss for log (total number of counts across the entire 1kb input sequence)
  - Multinomial loss for the likelihood of the observed count distribution over the sequence, compared to the predicted probabilities



Stranded profile shape + signal amplitude

$$Loss = -\log p_{mult.}(\mathbf{k}^{obs} \mid \mathbf{p}^{pred}, n^{obs}) + \lambda(\log(1+n^{obs}) - \log(1+n^{pred}))^2$$

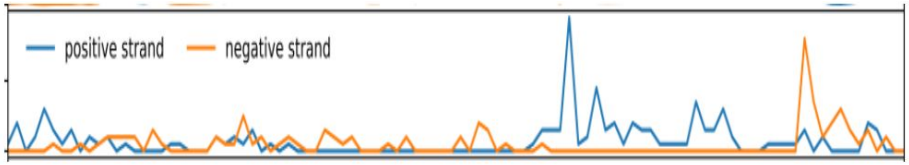
k^{obs}: vector of observed reads counts at each position
p^{pred}: learned multinomial prob. at each position
n^{obs}: total number of read counts across entire 1 kb

Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.

Slide Credit: Anshul Kundaje

BPNet: Profile regression loss

- Two-part loss function for optimizing prediction of the binding profile across the input sequence
  - MSE loss for log (total number of counts across the entire 1kb input sequence)
  - Multinomial loss for the likelihood of the observed count distribution over the sequence, compared to the predicted probabilities



Stranded profile shape + signal amplitude

$$Loss = -\log p_{mult.}(\mathbf{k}^{obs} \mid \mathbf{p}^{pred}, n^{obs}) + \lambda(\log(1+n^{obs}) - \log(1+n^{pred}))^2$$

k^{obs}: vector of observed reads counts at each position
p^{pred}: learned multinomial prob. at each position
n^{obs}: total number of read counts across entire 1 kb

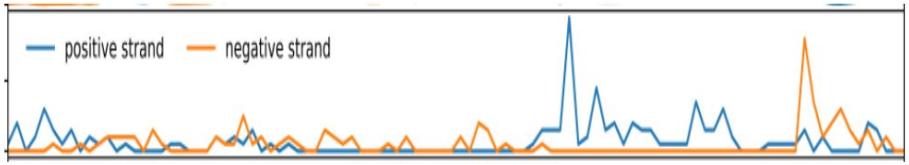
MSE loss

Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.

Slide Credit: Anshul Kundaje

BPNet: Profile regression loss

- Two-part loss function for optimizing prediction of the binding profile across the input sequence
  - MSE loss for log (total number of counts across the entire 1kb input sequence)
  - Multinomial loss for the likelihood of the observed count distribution over the sequence, compared to the predicted probabilities



Stranded profile shape + signal amplitude

$$Loss = -\log p_{mult.}(\mathbf{k}^{obs} \mid \mathbf{p}^{pred}, n^{obs}) + \lambda(\log(1+n^{obs}) - \log(1+n^{pred}))^2$$

k^{obs}: vector of observed reads counts at each position
p^{pred}: learned multinomial prob. at each position
n^{obs}: total number of read counts across entire 1 kb

Multinomial loss

Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.

Slide Credit: Anshul Kundaje

Multinomial loss component

$$Loss = -\log p_{mult.}(\mathbf{k}^{obs} \mid \mathbf{p}^{pred}, n^{obs}) + \lambda(\log(1+n^{obs}) - \log(1+n^{pred}))^2$$

k^{obs}: vector of observed reads counts at each position
p^{pred}: learned multinomial prob. at each position
n^{obs}: total number of read counts across entire 1 kb

Multinomial loss

Multinomial probability distribution

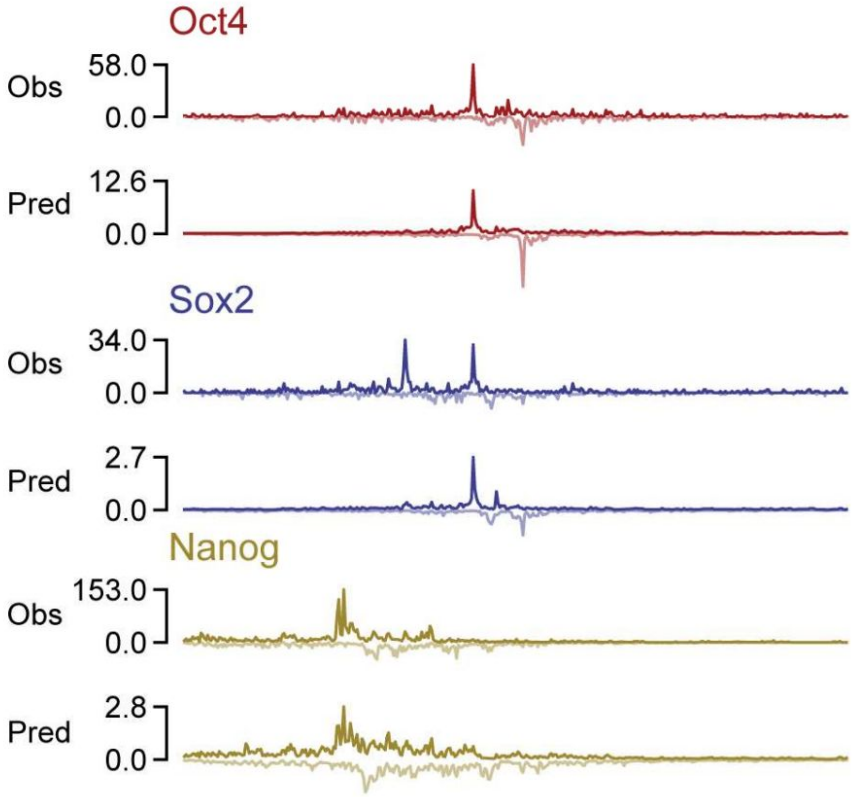
Suppose one does an experiment of extracting n^{obs} balls of 1000 different colors from a bag. Denote as p_i the probability that a given extraction will be in color *i*. Let k_i be the number of balls extracted of color i. The probability of this multinomial distribution is

$$p_{mult}([k_1, k_2 \ldots k_{1000}] \mid [p_1, p_2, \ldots, p_{1000}], n^{obs}) = \frac{n^{obs}!}{k_1!\, k_2! \ldots k_{1000}!} p_1^{k_1} p_2^{k_2} \ldots p_{1000}^{k_{1000}}$$

Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.

Slide Credit: Anshul Kundaje

BPNet predicted TF profiles



Avsec et al. Deep learning at base-resolution reveals motif syntax of the cis-regulatory code, 2019.

Slide Credit: Anshul Kundaje

Next topic: Multimodal data

Multimodal data

Can be very similar, e.g. different image acquisition variants

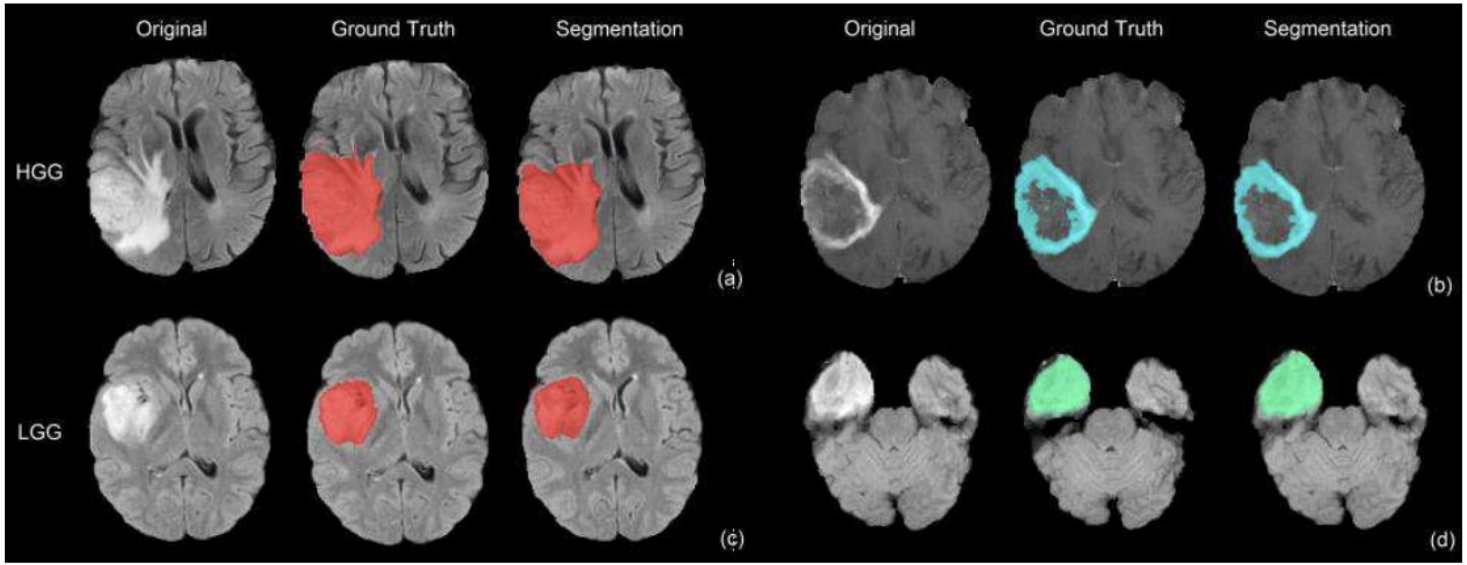


Figure credit: Dong et al. MIUA, 2017.

Multimodal data

Or very different, e.g. different types of clinical data

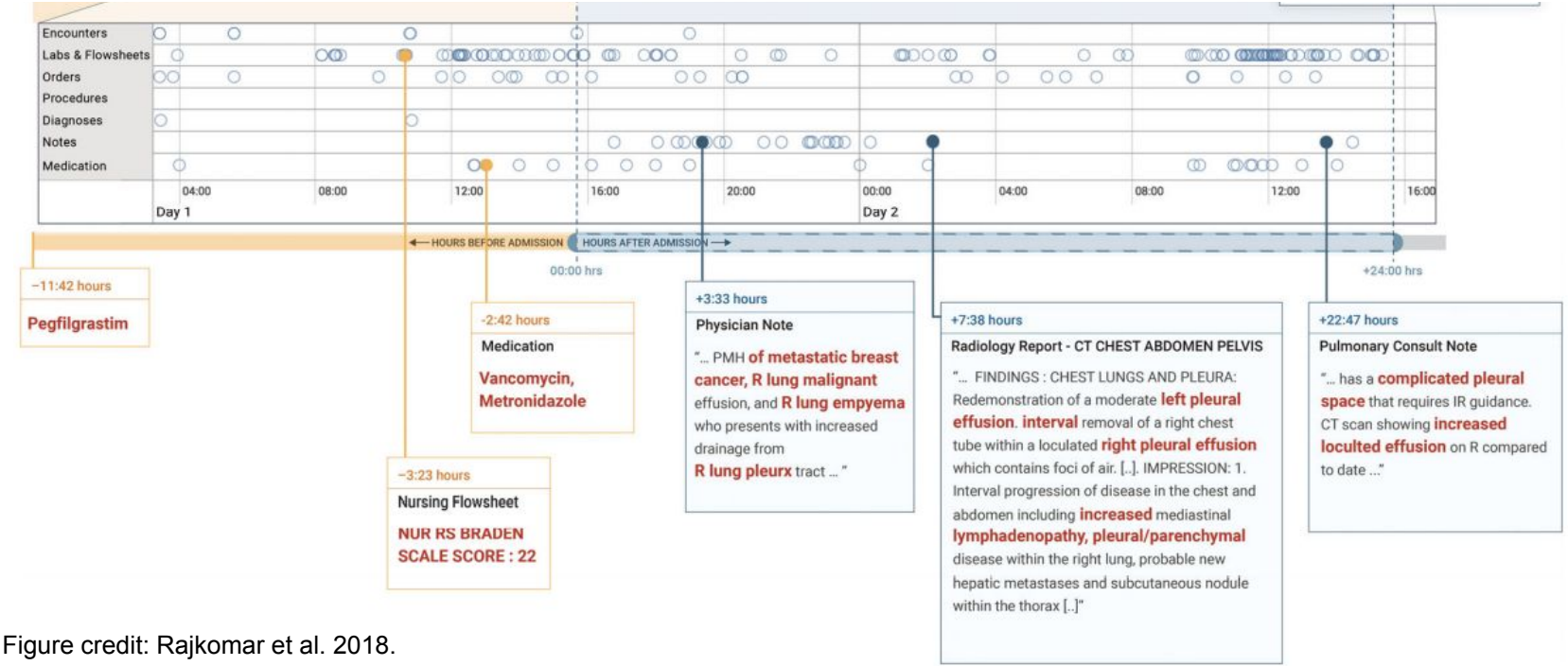
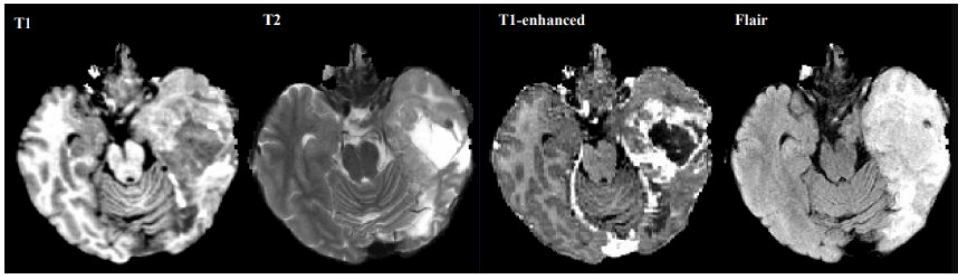


Figure credit: Rajkomar et al. 2018.

Similar data: can fuse at input

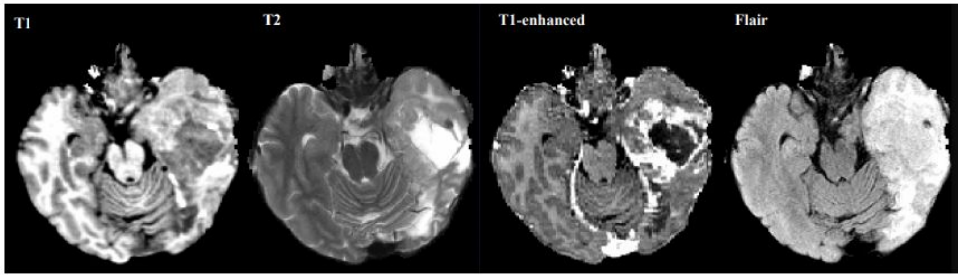
- Havaei et al.: brain tumor segmentation from multimodal MR images



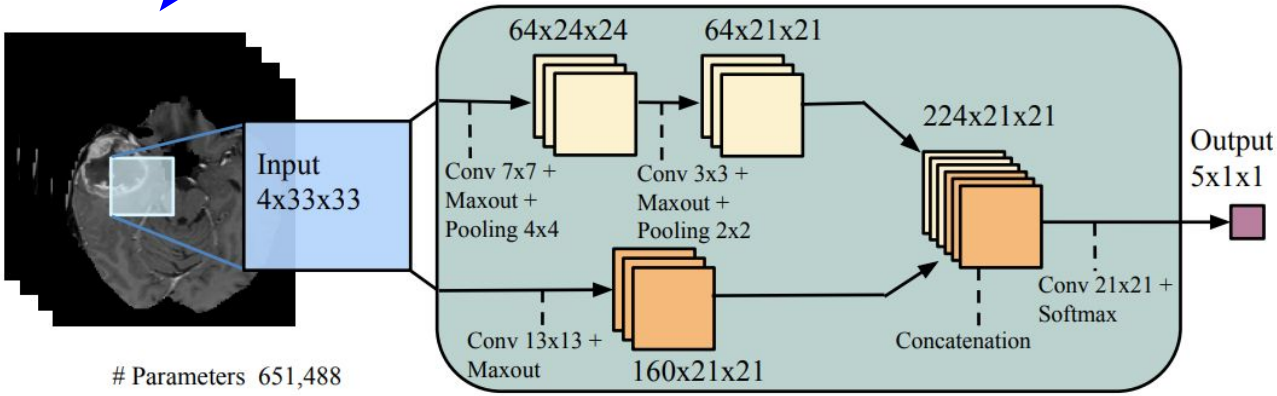
Havaei et al. Brain Tumor Segmentation with Deep Neural Networks. Medical Image Analysis, 2016.

Similar data: can fuse at input

- Havaei et al.: brain tumor segmentation from multimodal MR images



Stack modalities such that each channel of input is a different modality.

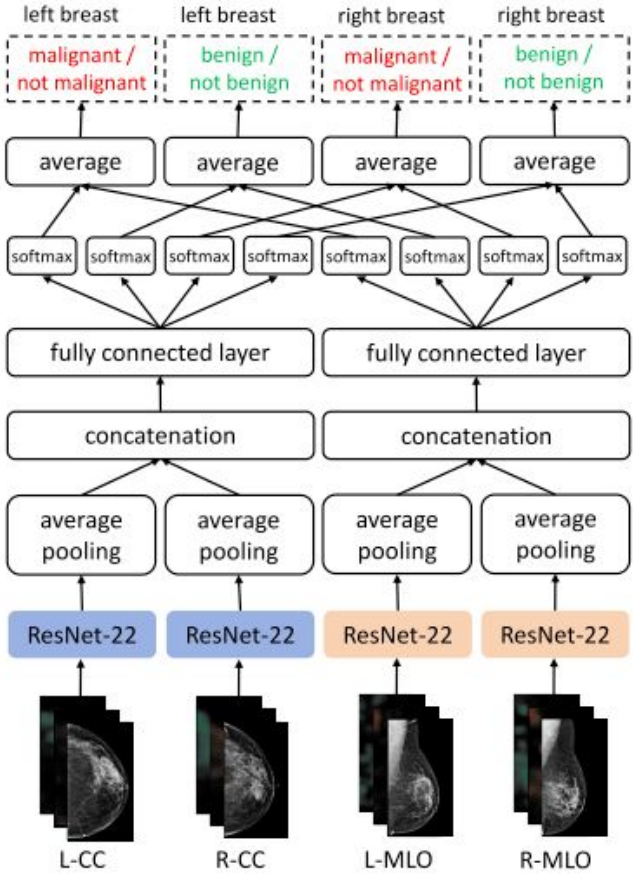


Havaei et al. Brain Tumor Segmentation with Deep Neural Networks. Medical Image Analysis, 2016.

More different data: may want some layers of modality-specific processing

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)



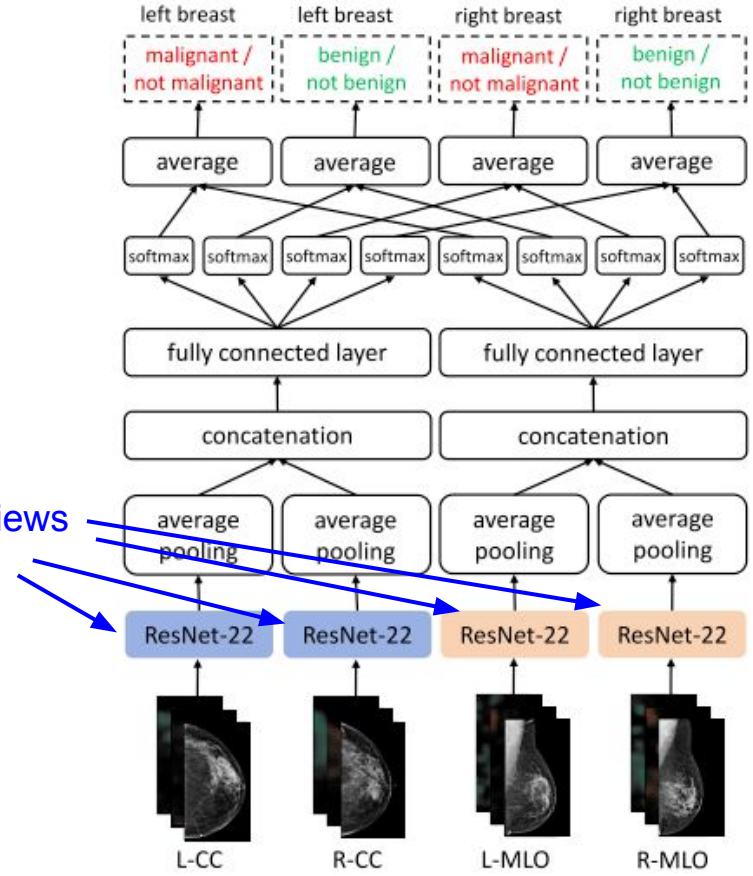
Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

More different data: may want some layers of modality-specific processing

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)

Separate initial processing for different mammogram views



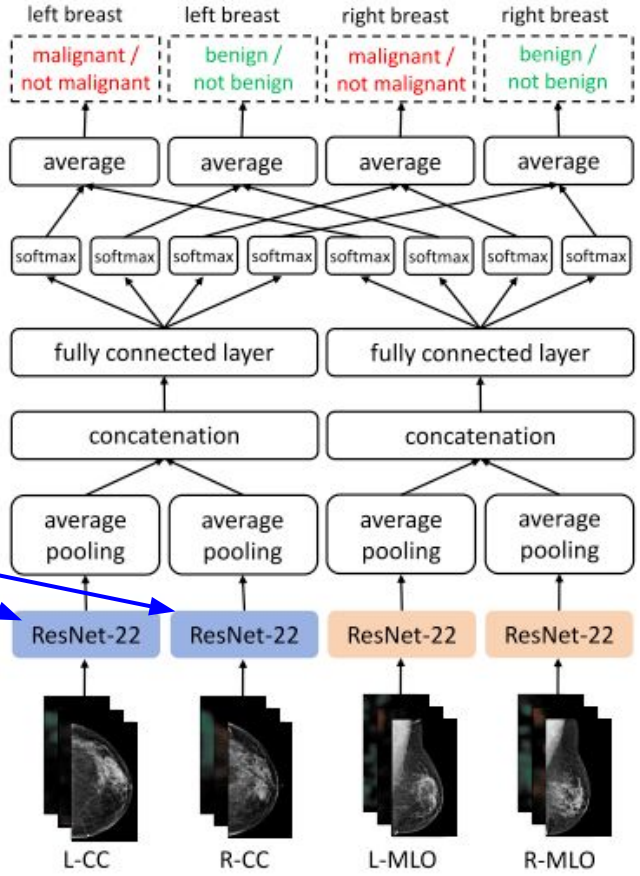
Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

More different data: may want some layers of modality-specific processing

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)

Shared weights across the two networks



Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

More different data: may want some layers of modality-specific processing

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)

More different views have separately learned parameters

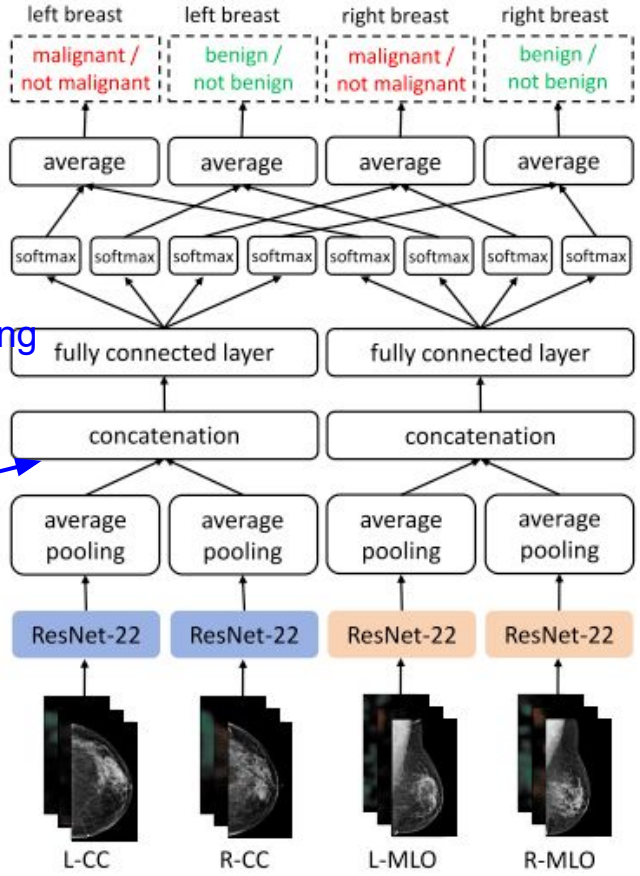
Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

More different data: may want some layers of modality-specific processing

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)

Multimodal fusion at intermediate part of processing (very common): concatenate outputs of modality-specific processing into one feature vector.



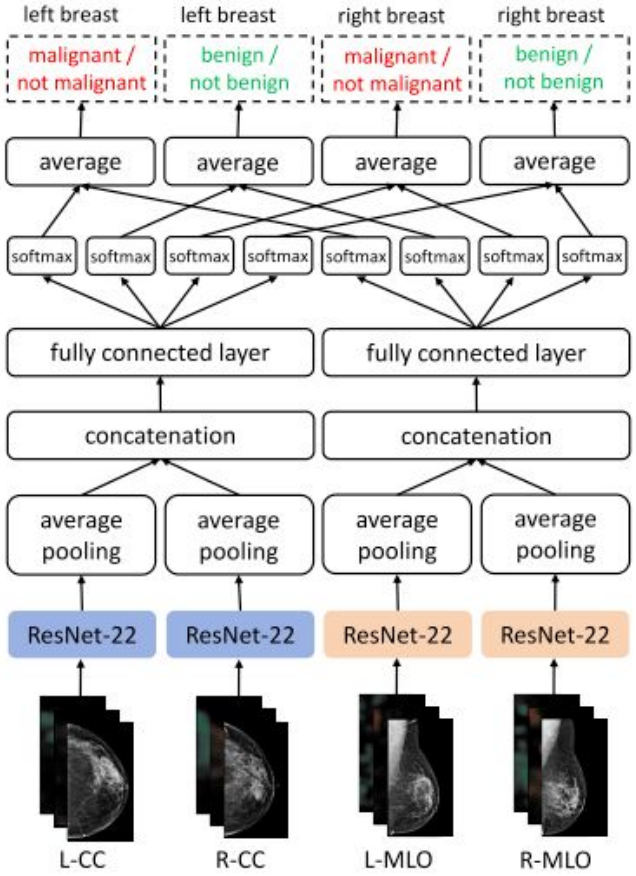
Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

More different data: may want some layers of modality-specific processing

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)

Fully connected layer (or several) afterwards. Concatenated feature vector no longer contains spatial relationships suitable for conv layers.



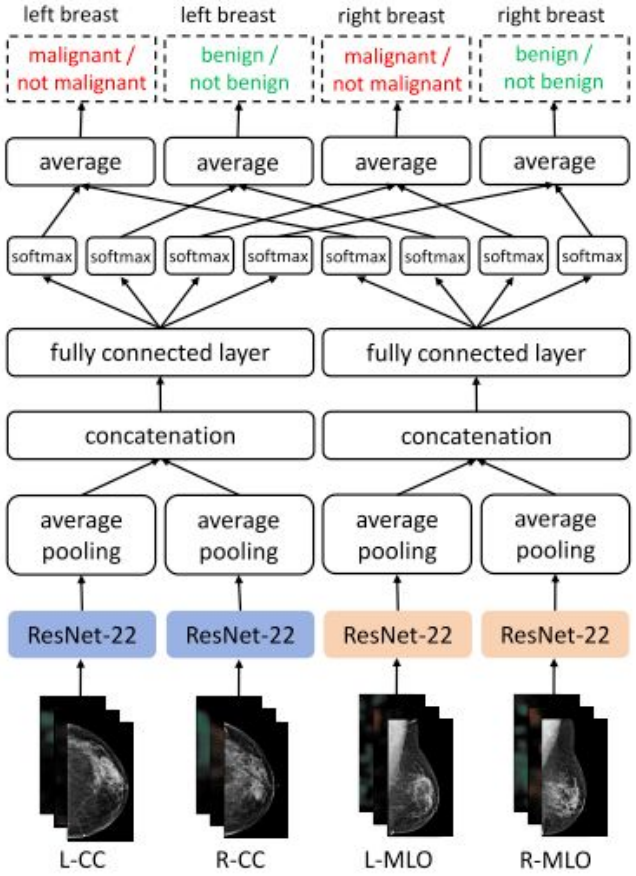
Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

More different data: may want some layers of modality-specific processing

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)

Predict all 4 binary outputs from each view



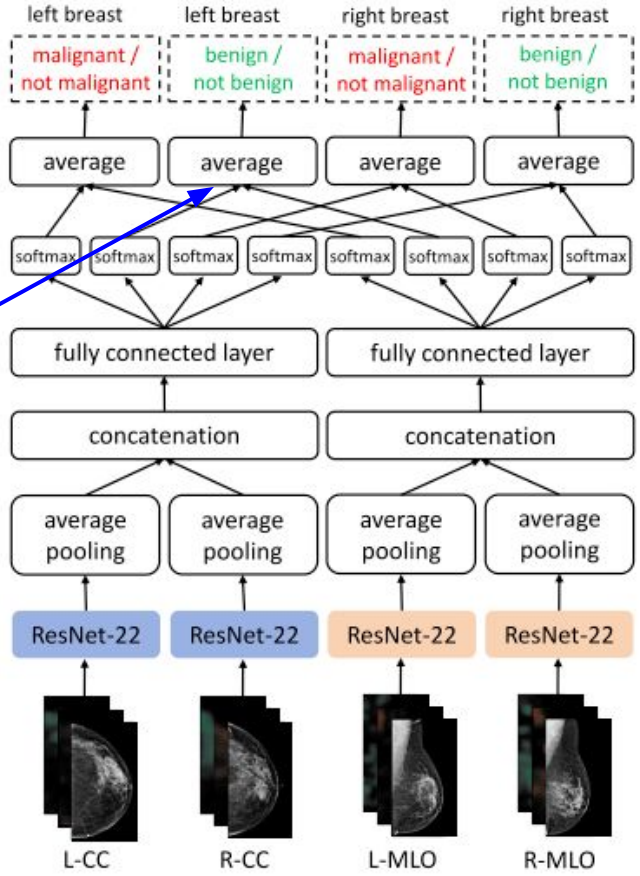
Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

More different data: may want some layers of modality-specific processing

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)

This model also uses a second type of fusion for the CC vs. MLO views: late fusion of predictions through averaging.

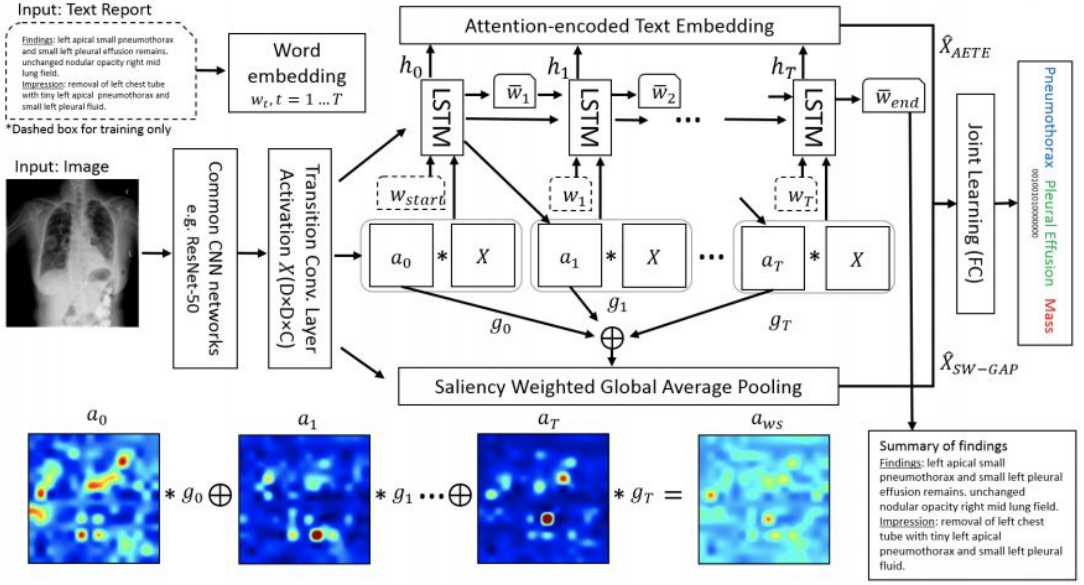


Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

A recurrent network approach for combining multimodal data

Wang et al. 2018:

- Jointly process chest x-rays and associated reports to produce disease labels that can be used to produce auto-annotation disease labels



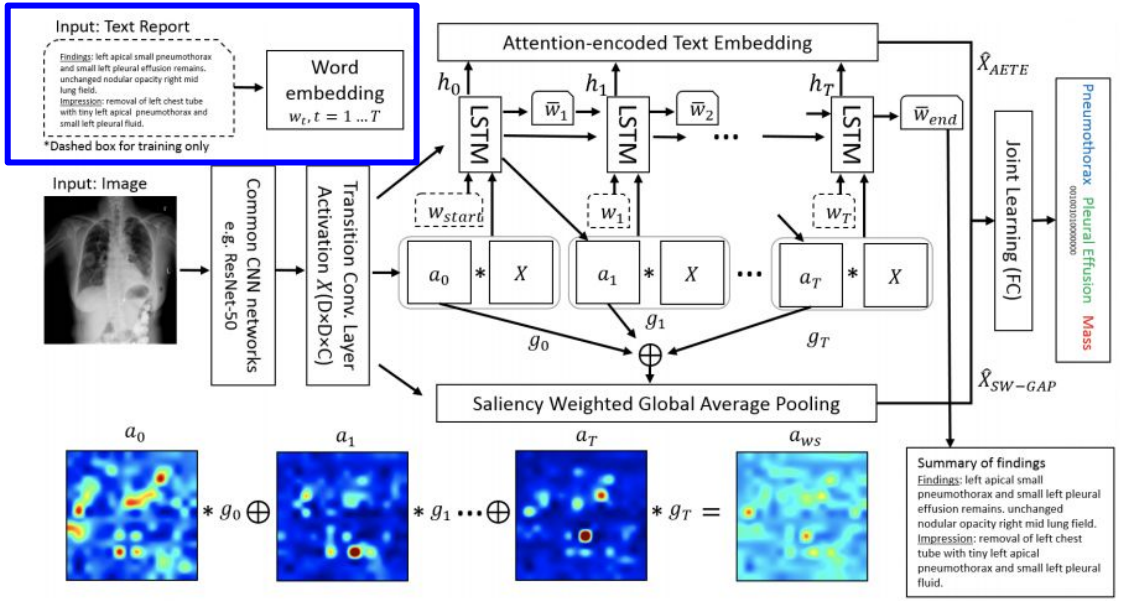
Wang et al. TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays. CVPR, 2018.

A recurrent network approach for combining multimodal data

Wang et al. 2018:

- Jointly process chest x-rays and associated reports to produce disease labels that can be used to produce auto-annotation disease labels

Use NLP approaches to generate word embedding representations of words in text



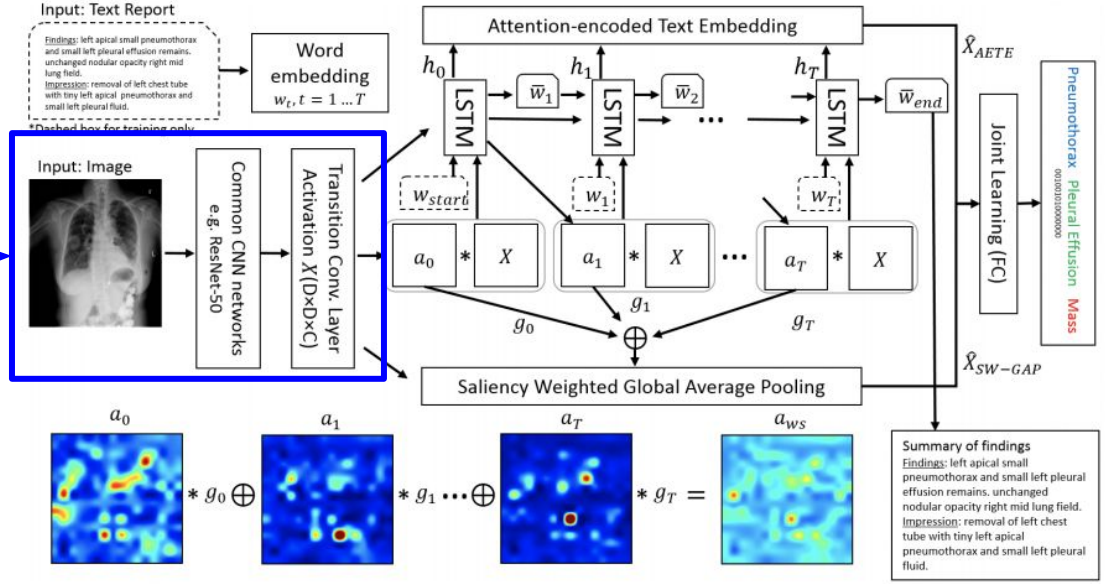
Wang et al. TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays. CVPR, 2018.

A recurrent network approach for combining multimodal data

Wang et al. 2018:

- Jointly process chest x-rays and associated reports to produce disease labels that can be used to produce auto-annotation disease labels

Use common CNN networks to generate feature representation of image data



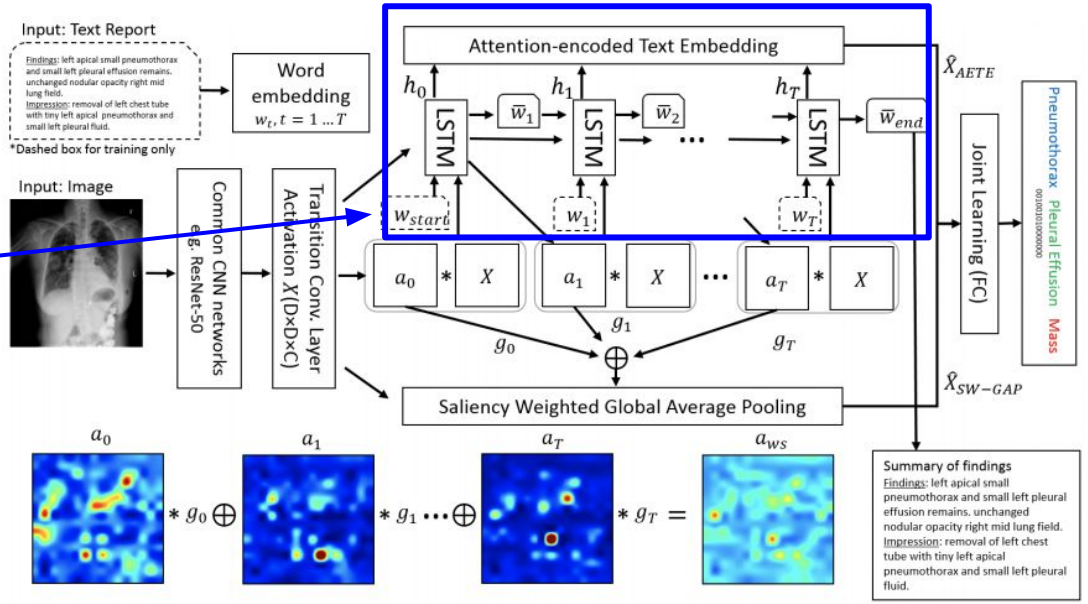
Wang et al. TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays. CVPR, 2018.

A recurrent network approach for combining multimodal data

Wang et al. 2018:

- Jointly process chest x-rays and associated reports to produce disease labels that can be used to produce auto-annotation disease labels

Use LSTM to process sequence of text data embedding representations



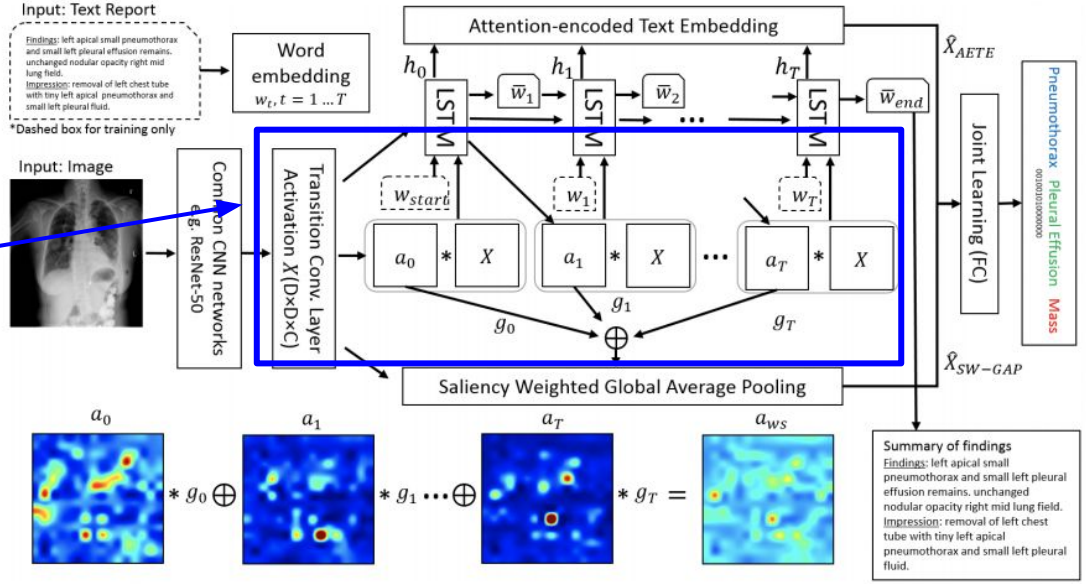
Wang et al. TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays. CVPR, 2018.

A recurrent network approach for combining multimodal data

Wang et al. 2018:

- Jointly process chest x-rays and associated reports to produce disease labels that can be used to produce auto-annotation disease labels

Image data is an additional input to the LSTM at each time step (with soft-attention weighting)



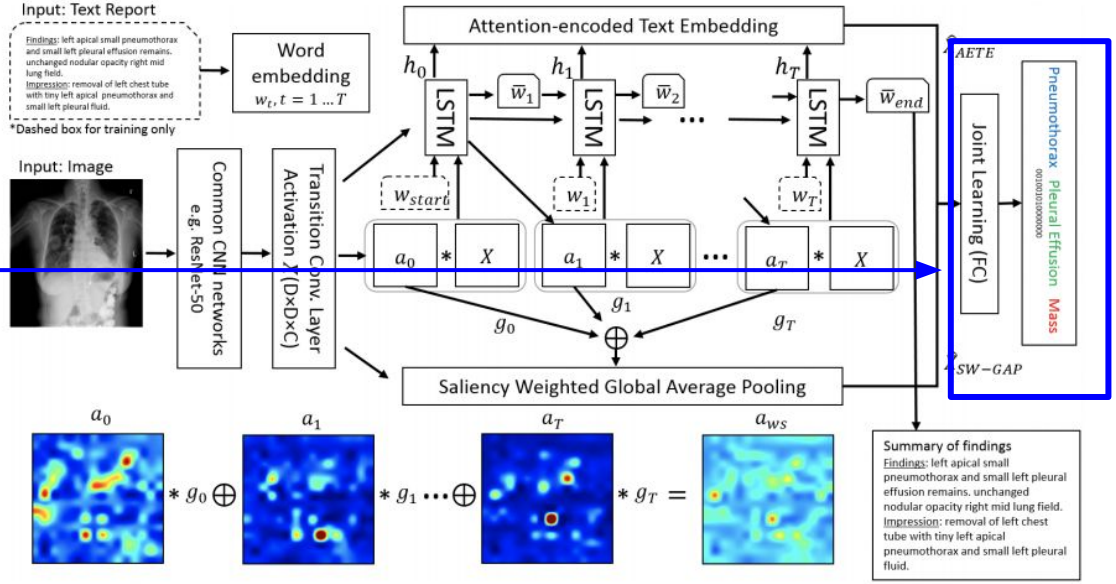
Wang et al. TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays. CVPR, 2018.

A recurrent network approach for combining multimodal data

Wang et al. 2018:

- Jointly process chest x-rays and associated reports to produce disease labels that can be used to produce auto-annotation disease labels

Final fully-connected layer fusion and prediction of disease labels

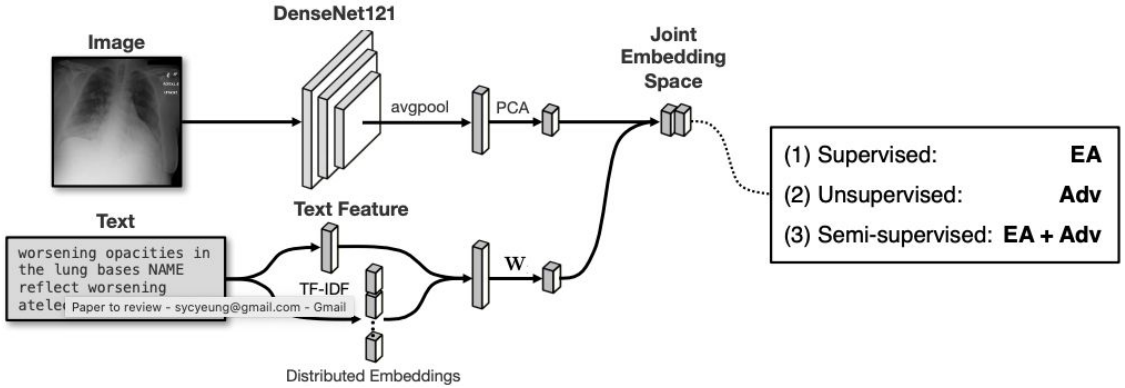


Wang et al. TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays. CVPR, 2018.

Another direction of research: learning multimodal embedding spaces

Hsu et al. 2018:

- Learn mapping from images and text to vectors in the same embedding space, such that images are embedded closer to their corresponding reports than other reports, and vice versa.
- Can be used for e.g. cross-domain retrieval

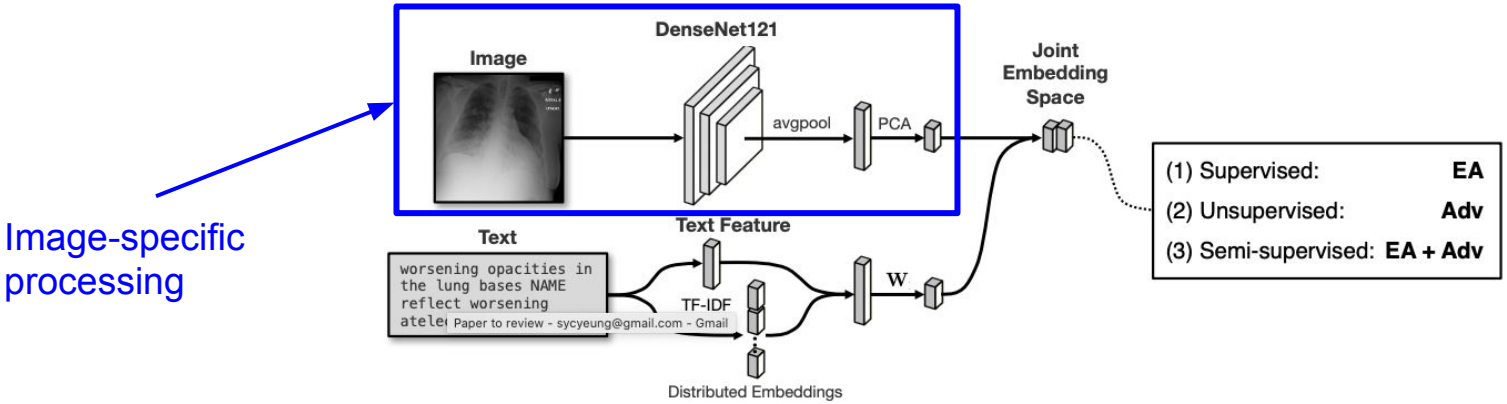


Hsu et al. Unsupervised Multimodal Representation Learning across Medical Images and Reports. NeurIPS ML4H, 2018.

Another direction of research: learning multimodal embedding spaces

Hsu et al. 2018:

- Learn mapping from images and text to vectors in the same embedding space, such that images are embedded closer to their corresponding reports than other reports, and vice versa.
- Can be used for e.g. cross-domain retrieval

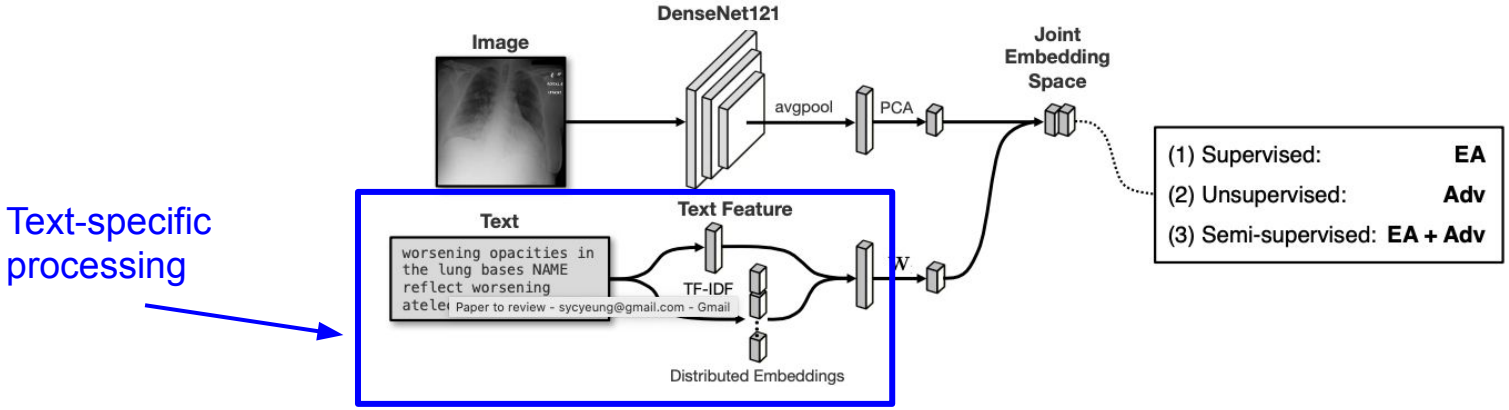


Hsu et al. Unsupervised Multimodal Representation Learning across Medical Images and Reports. NeurIPS ML4H, 2018.

Another direction of research: learning multimodal embedding spaces

Hsu et al. 2018:

- Learn mapping from images and text to vectors in the same embedding space, such that images are embedded closer to their corresponding reports than other reports, and vice versa.
- Can be used for e.g. cross-domain retrieval



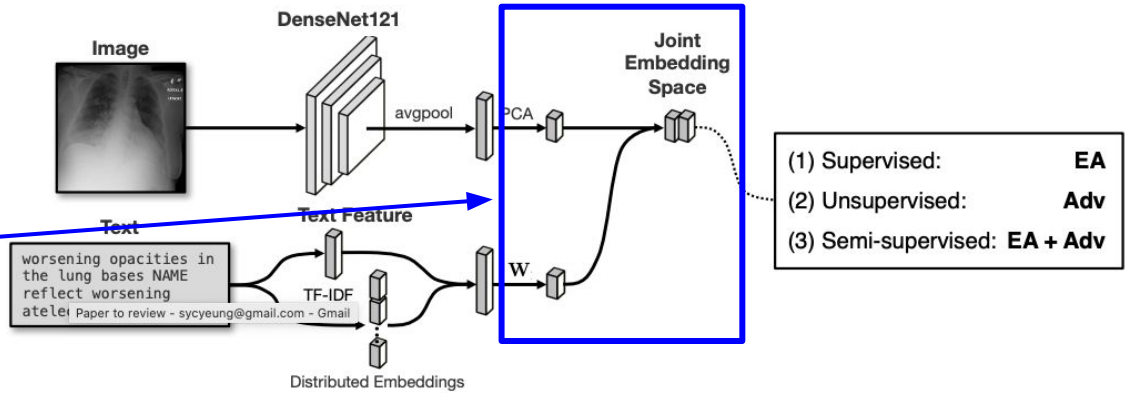
Hsu et al. Unsupervised Multimodal Representation Learning across Medical Images and Reports. NeurIPS ML4H, 2018.

Another direction of research: learning multimodal embedding spaces

Hsu et al. 2018:

- Learn mapping from images and text to vectors in the same embedding space, such that images are embedded closer to their corresponding reports than other reports, and vice versa.
- Can be used for e.g. cross-domain retrieval

Mapping to joint embedding space



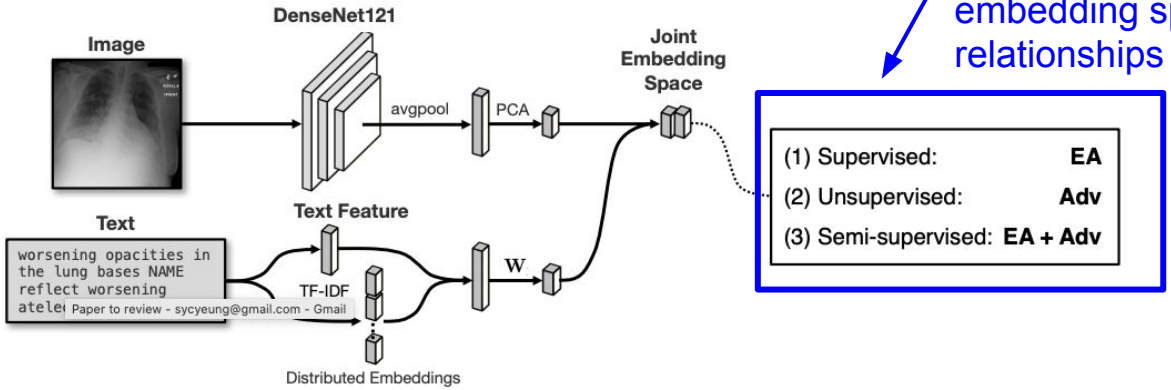
Hsu et al. Unsupervised Multimodal Representation Learning across Medical Images and Reports. NeurIPS ML4H, 2018.

Another direction of research: learning multimodal embedding spaces

Hsu et al. 2018:

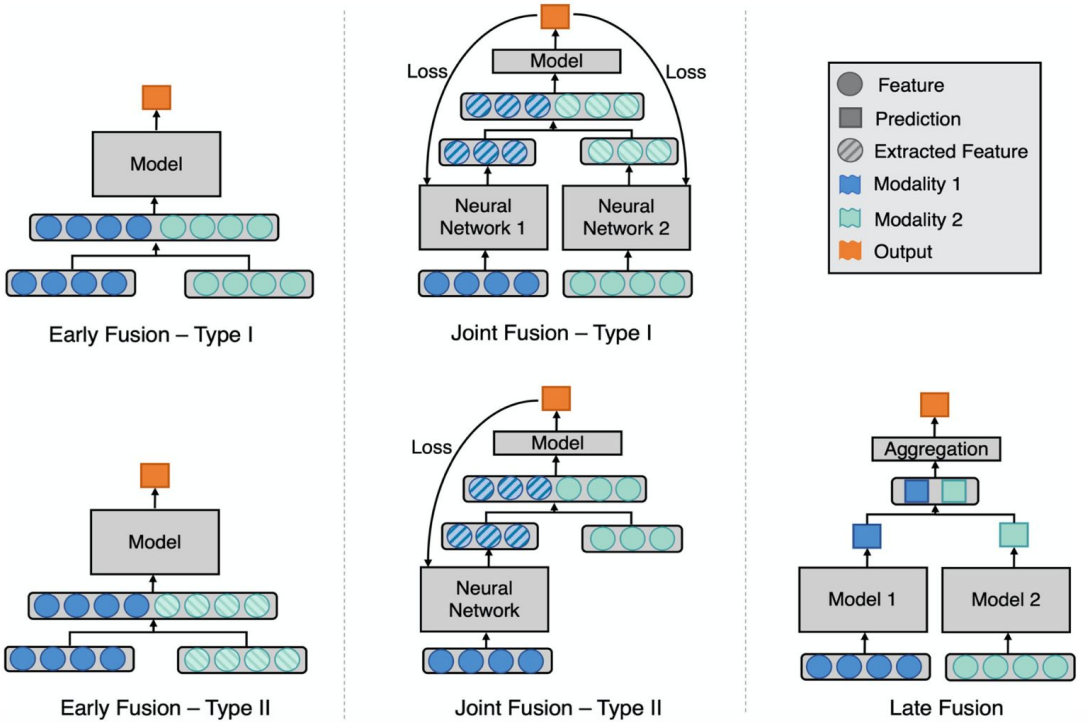
- Learn mapping from images and text to vectors in the same embedding space, such that images are embedded closer to their corresponding reports than other reports, and vice versa.
- Can be used for e.g. cross-domain retrieval

Different loss objectives can be used to encourage desired embedding space relationships



Hsu et al. Unsupervised Multimodal Representation Learning across Medical Images and Reports. NeurIPS ML4H, 2018.

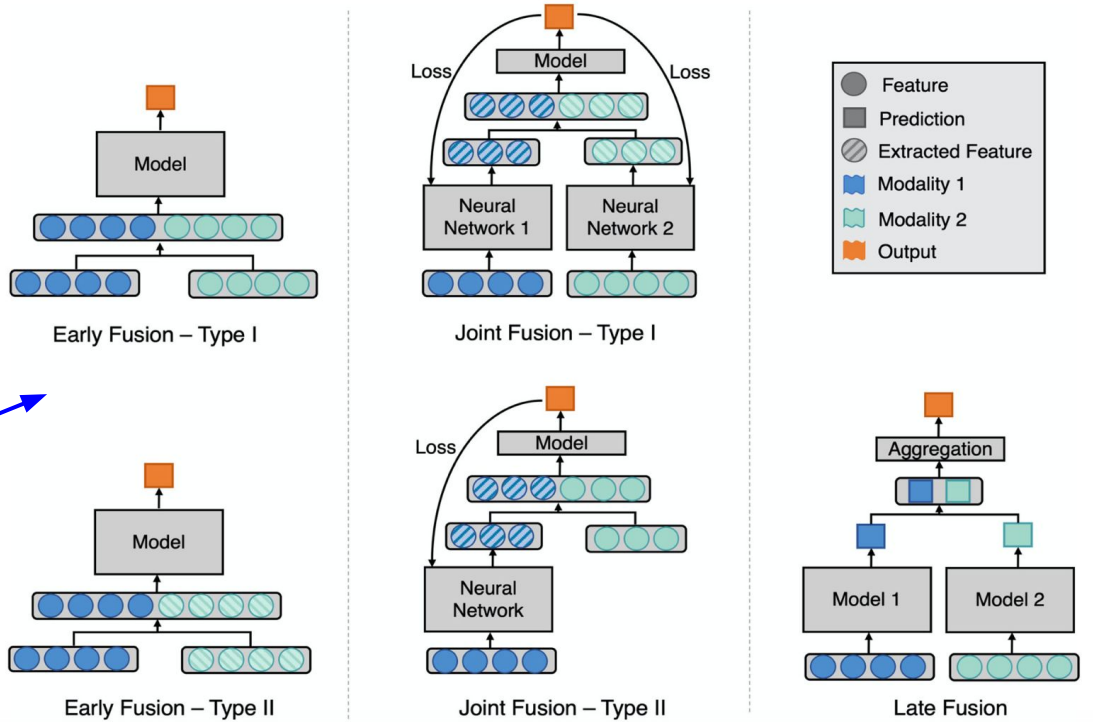
Categorizations of multimodal models



Huang et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines, 2020.

Categorizations of multimodal models

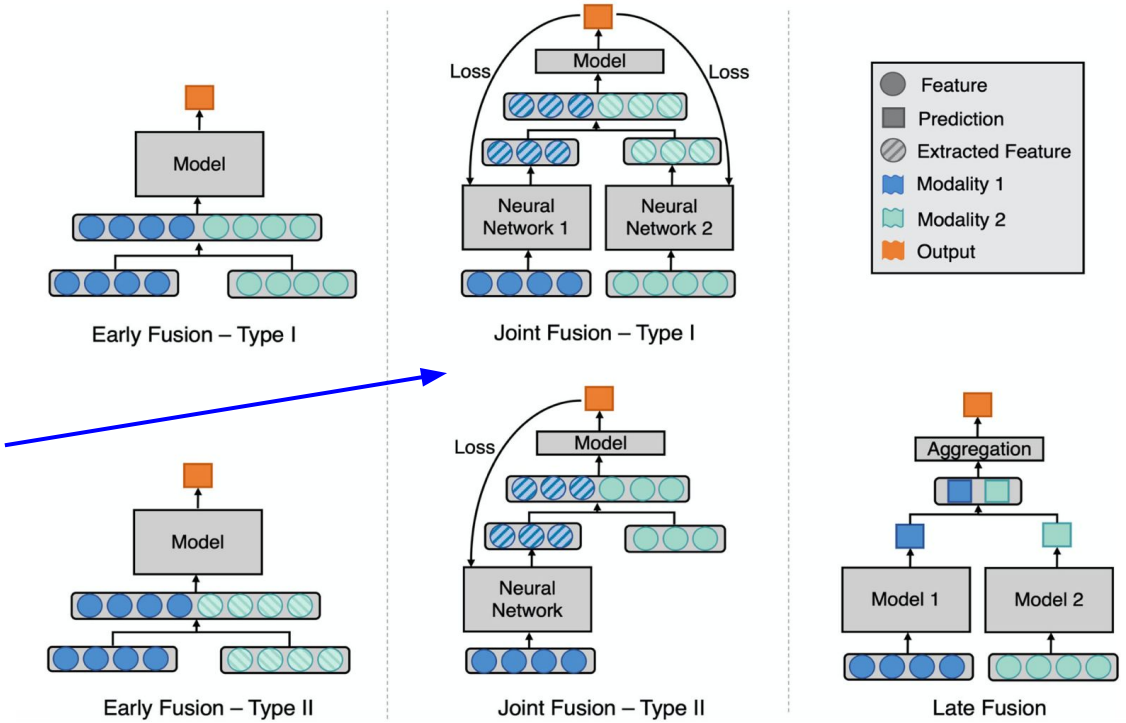
Early fusion: concatenate / combine data before any model processing. Includes using extracted features as input, if model gradients are not backpropagated to update feature extractor parameters



Huang et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines, 2020.

Categorizations of multimodal models

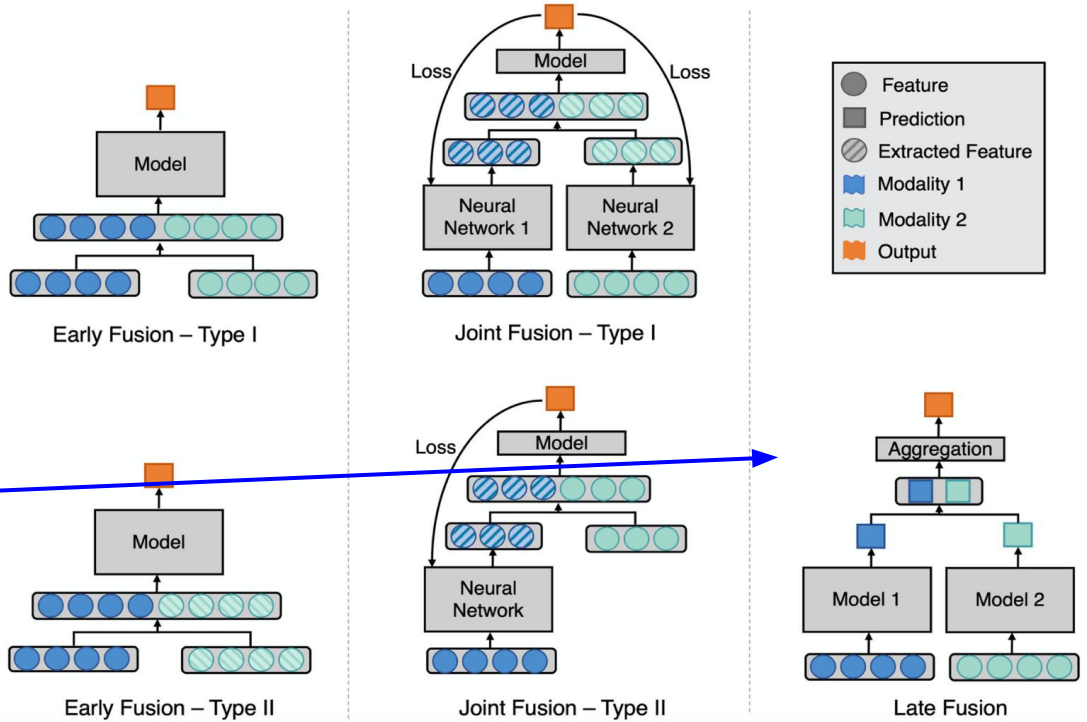
Joint fusion: Both modality-specific components (with learnable parameters) and combined-modality components within the model, that are updated during model training



Huang et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines, 2020.

Categorizations of multimodal models

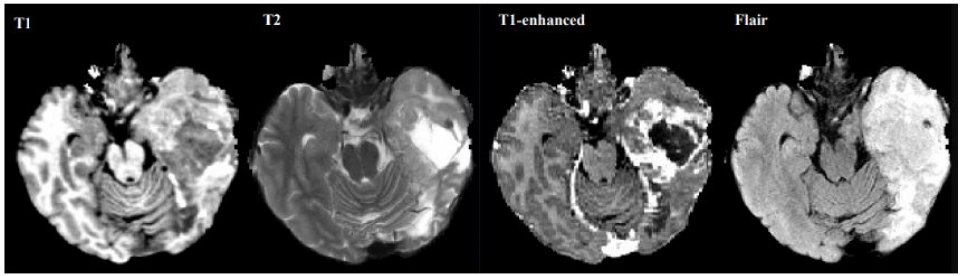
Late fusion: Main learnable model components are only model specific. Individual modality outputs are then aggregated.

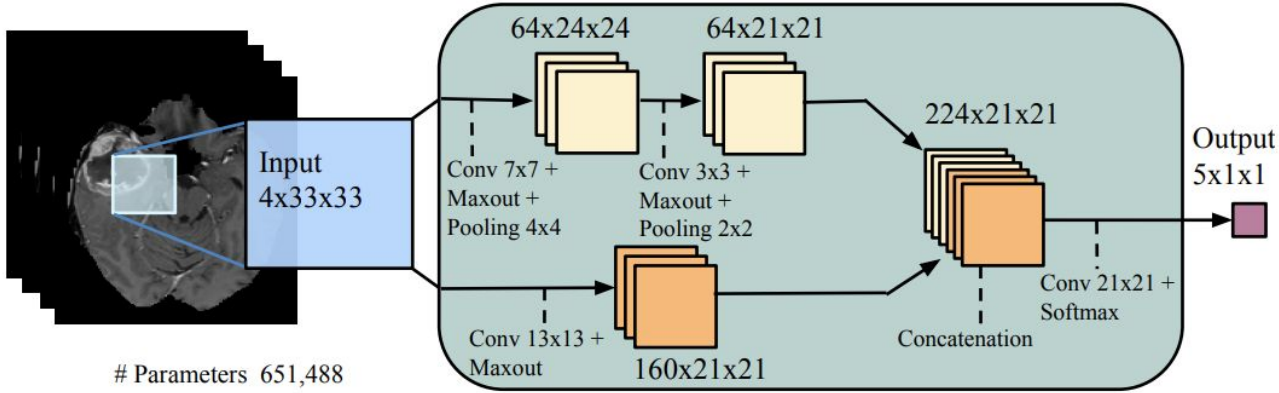


Huang et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines, 2020.

Q: What kind of fusion was this model?

- Havaei et al.: brain tumor segmentation from multimodal MR images



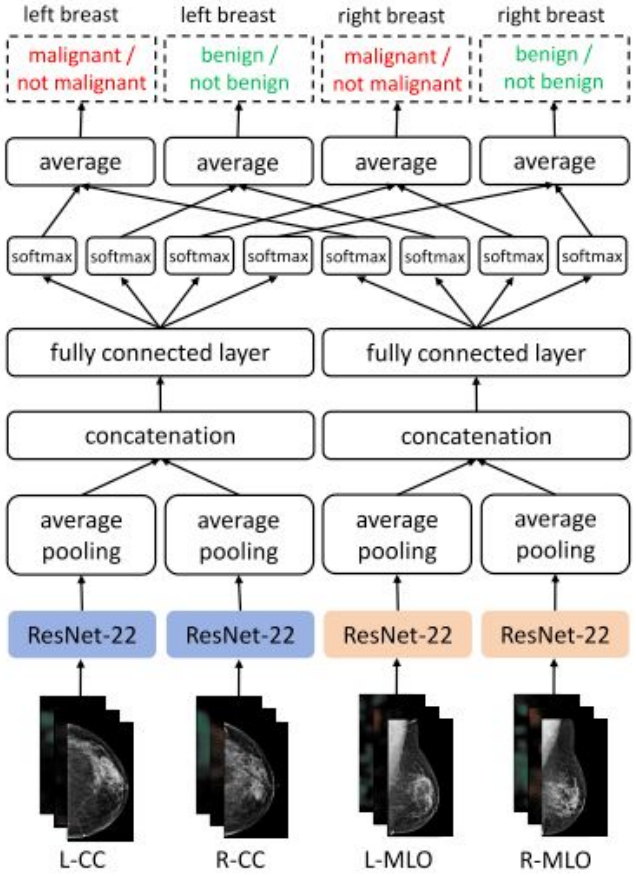


Havaei et al. Brain Tumor Segmentation with Deep Neural Networks. Medical Image Analysis, 2016.

Q: What kind of fusion was this model?

Wu et al. 2019:

- Binary classification of breast malignant and benign findings
- Model based on ResNet architecture
- Multi-view network (different views can be considered different modalities)

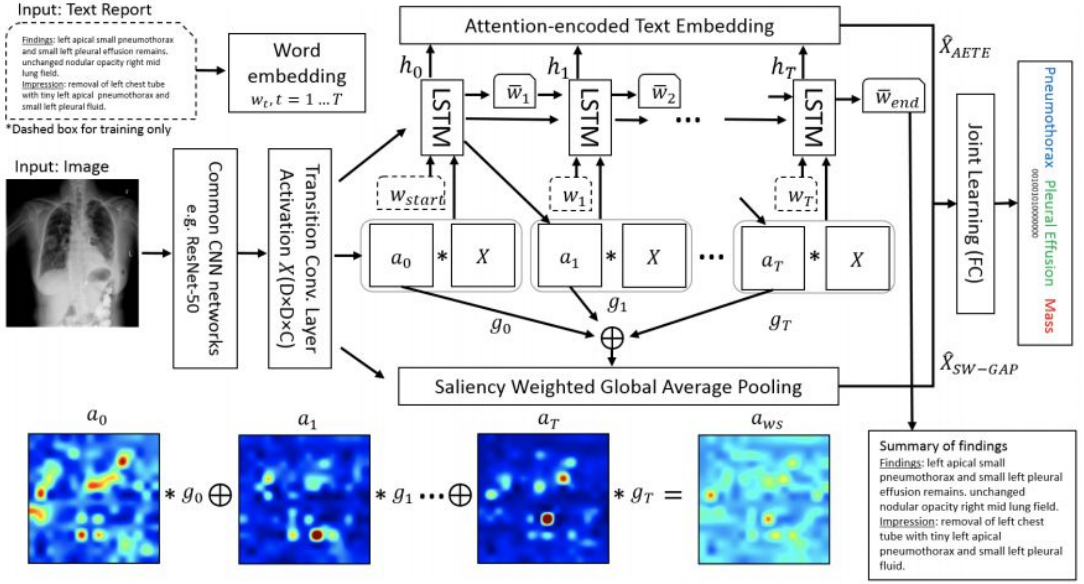


Wu et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Trans Med Imaging, 2019.

Q: What kind of fusion was this model?

Wang et al. 2018:

- Jointly process chest x-rays and associated reports to produce disease labels that can be used to produce auto-annotation disease labels



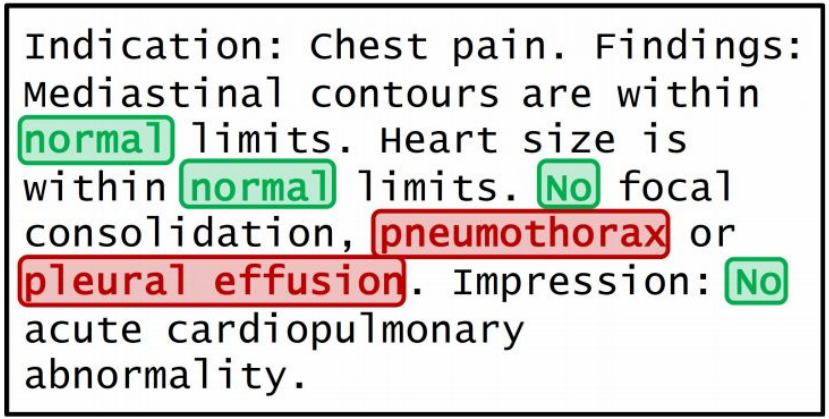
Wang et al. TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays. CVPR, 2018.

Weak Supervision

- Machine learning paradigm where labels for supervised training are obtained from noisy or imprecise (but more easily accessible) sources
- One possibility is through corresponding data available in a different modality! (e.g., radiology reports as a source of weak supervision for radiology images)

Weak supervision from radiology reports

Can use rule-based approaches for obtaining labels from free-text radiology reports



Normal Report

```
def LF_pneumothorax(c):
  if re.search(r'pneumo.*', c.report.text):
      return "ABNORMAL"

def LF_pleural_effusion(c):
  if "pleural effusion" in c.report.text:
      return "ABNORMAL"

def LF_normal_report(c, thresh=2):
    if len(NORMAL_TERMS.intersection(c.
     report.words)) > thresh:
      return "NORMAL"
```

LFs

Figure credit: Nishith Khandwala et al., 2017.
Dunmon et al. Cross-Modal Data Programming Enables Rapid Medical Machine Learning, 2020.

How can we produce good labels from noisy sources?

One approach: Aggregate multiple rules (labeling functions) with majority voting

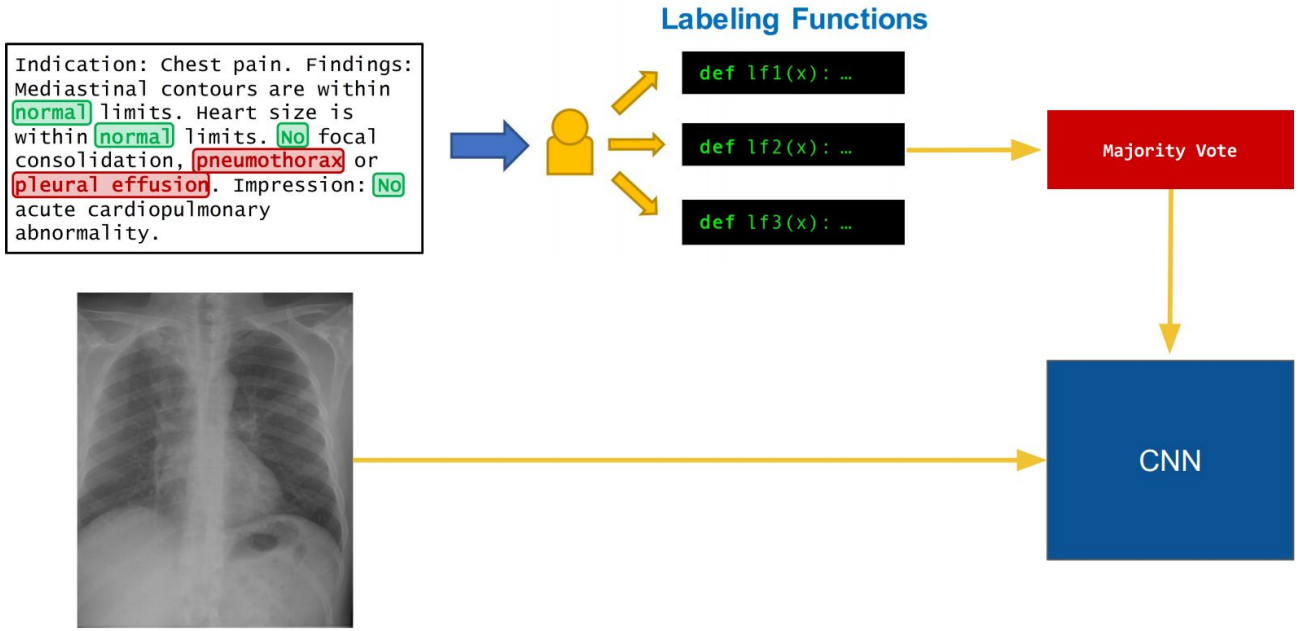


Figure credit: Nishith Khandwala et al., 2017.
Dunmon et al. Cross-Modal Data Programming Enables Rapid Medical Machine Learning, 2020.

How can we produce good labels from noisy sources?

More sophisticated approach: learn models for how to best aggregate noisy labeling functions!

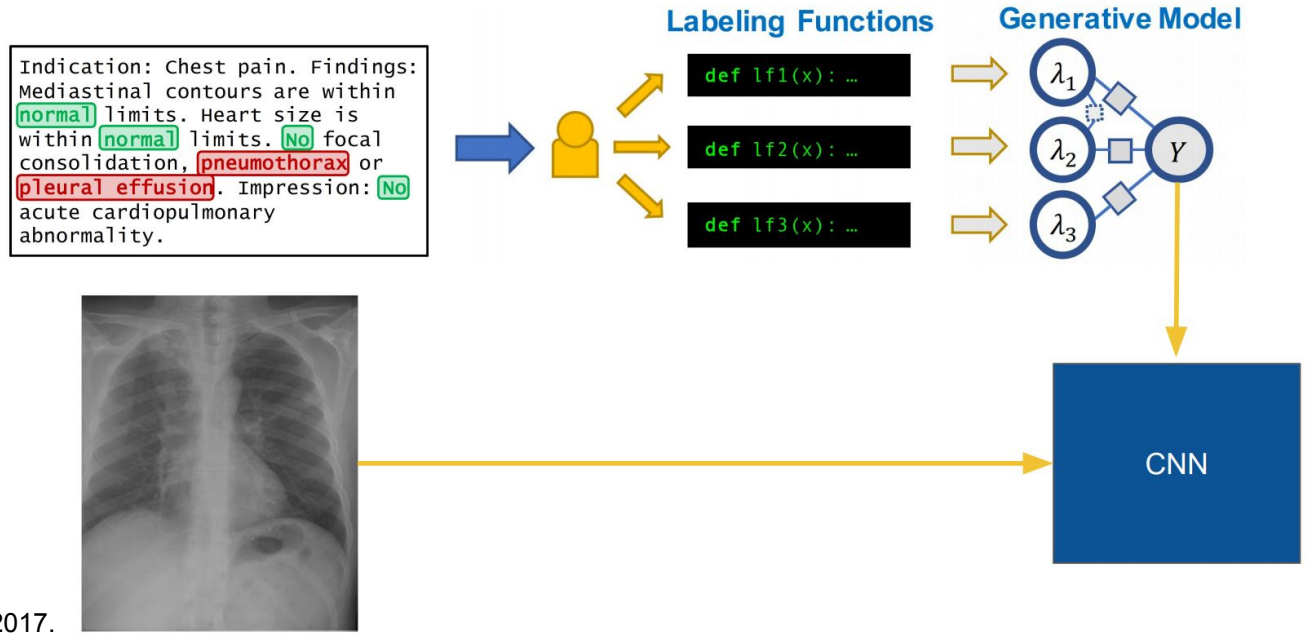


Figure credit: Nishith Khandwala et al., 2017.
Dunmon et al. Cross-Modal Data Programming Enables Rapid Medical Machine Learning, 2020.

How can we produce good labels from noisy sources?

More sophisticated approach: learn models for how to best aggregate noisy labeling functions!

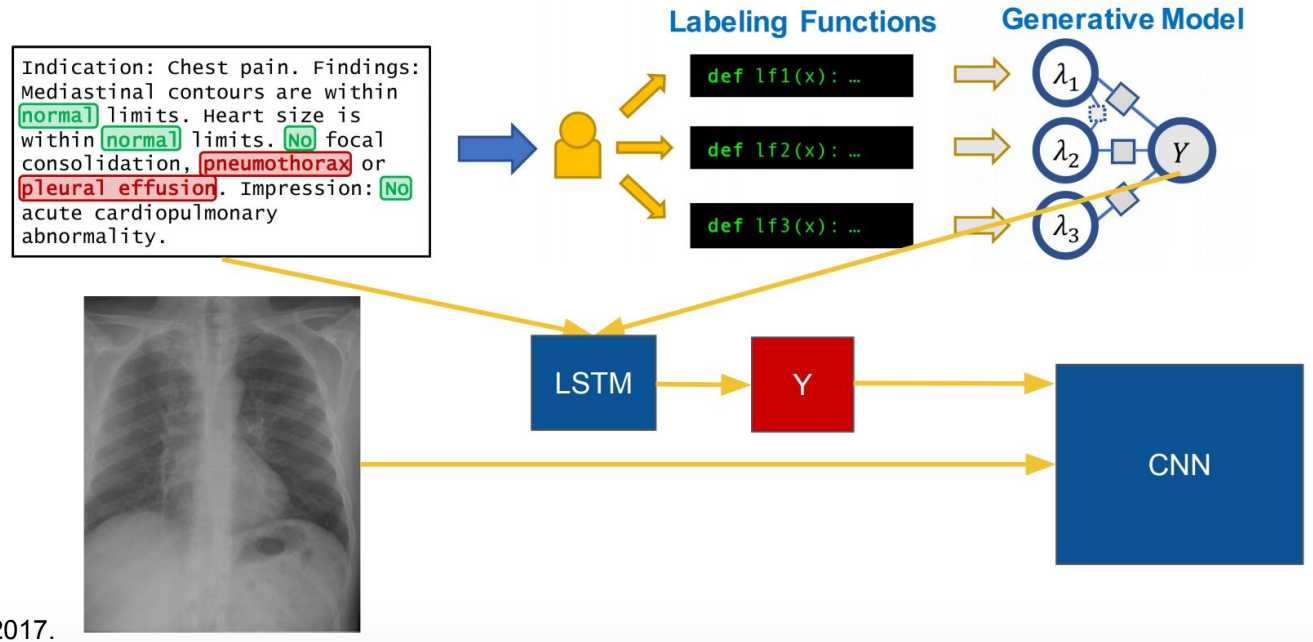
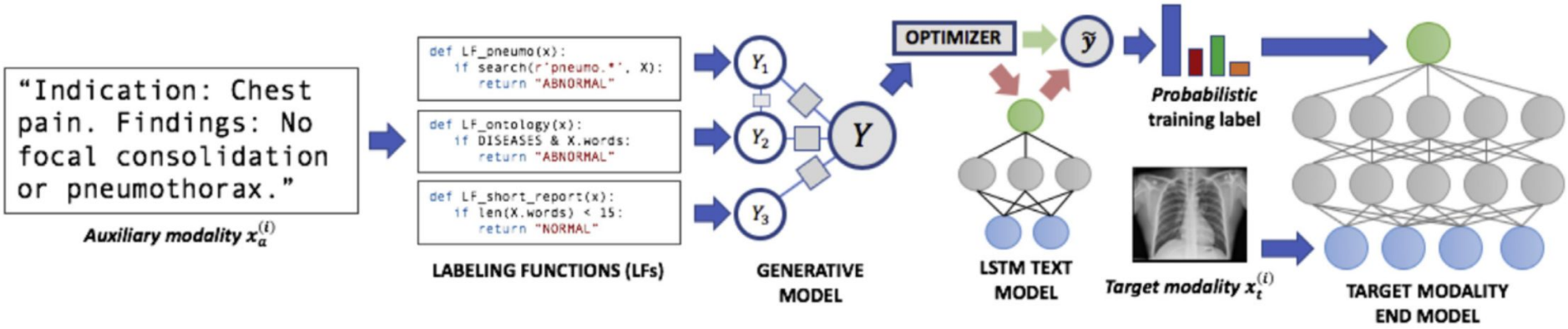


Figure credit: Nishith Khandwala et al., 2017.
Dunmon et al. Cross-Modal Data Programming Enables Rapid Medical Machine Learning, 2020.

“Data programming” paradigm for weak supervision



Dunmon et al. Cross-Modal Data Programming Enables Rapid Medical Machine Learning, 2020.

Summary

Today we covered:

- One more example of deep learning in genomics
- Multimodal data and models
- Weakly supervised learning

Next time:

- Special topics: AI for COVID-19