Data Science for Speech Therapy
Computer Vision and VFSS

Clinical sharing for continuous education

ST Benjamin Chow

Content

Data Science

- Recap
- Computer Vision
  - Image vs video
  - Reading images
  - Reading videos (image task)
  - Reading videos (action task)
  - Learning images (CNN)

Clinical Application

- Current practice
- Use cases
  1. hyoid bone detection
  2. pharyngeal phase
  3. PAS
  4. abnormal swallow

- Strategies to use CV

Computer Vision

- Data types: more than just spreadsheet

- Visual Data Source
  - Images
  - Videos/ gifs

Image vs Videos

	Image	Video
Spatial information *(where?)*	Yes	Yes
Temporal information *(when?)*	N.A.	Yes

http://dxline.info/img/new_ail/barium-swallow-and-meal_1.gif

Image vs Videos

	Image	Video
Spatial information *(where?)*	Yes	Yes
Common Tasks		Image tasks • Image Classification • Image Detection • Image Generation
Temporal information *(when?)*	N.A.	Yes
Unique Tasks	Nil	Action tasks • Action recognition

180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215

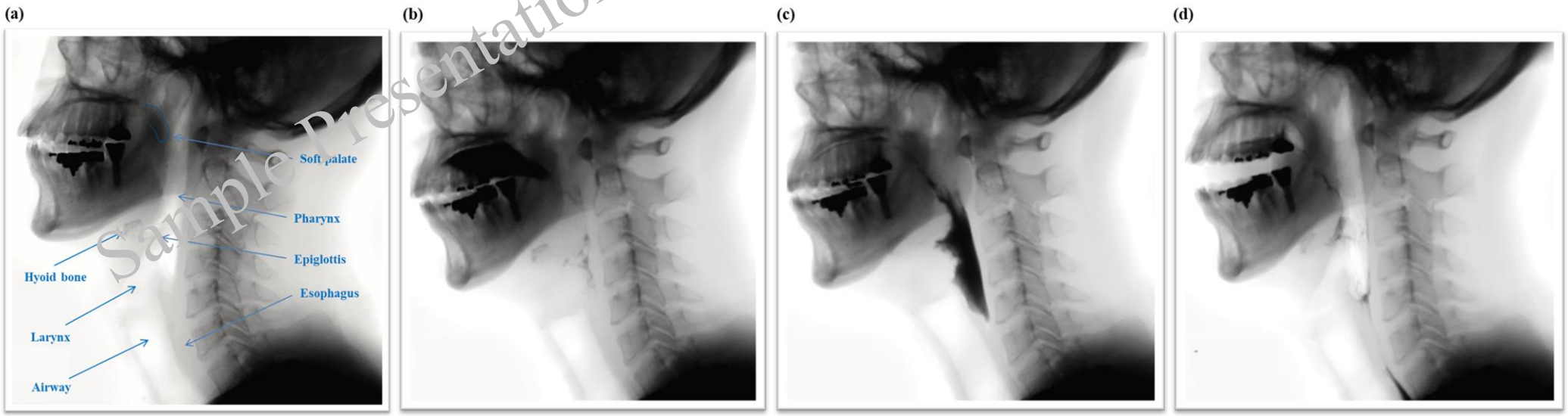
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215

- Image is made of multiple pixels
- Each pixel has an intensity value from 0 (absent colour) – 255

Importing Videos for image tasks

	Image	**Video**
Spatial information (where?)	*Yes*	**Yes**
Common Tasks		**Image tasks** • **Image Classification** • **Image Detection** • **Image Generation**
Temporal information (when?)	*N.A.*	*Yes*
Unique Tasks	*Nil*	*Action tasks* • *Action recognition*

- Multiple image frames in sequence.
- Extract images frames from videos at fixed time intervals.



https://www.mdpi.com/1424-8220/19/18/3873

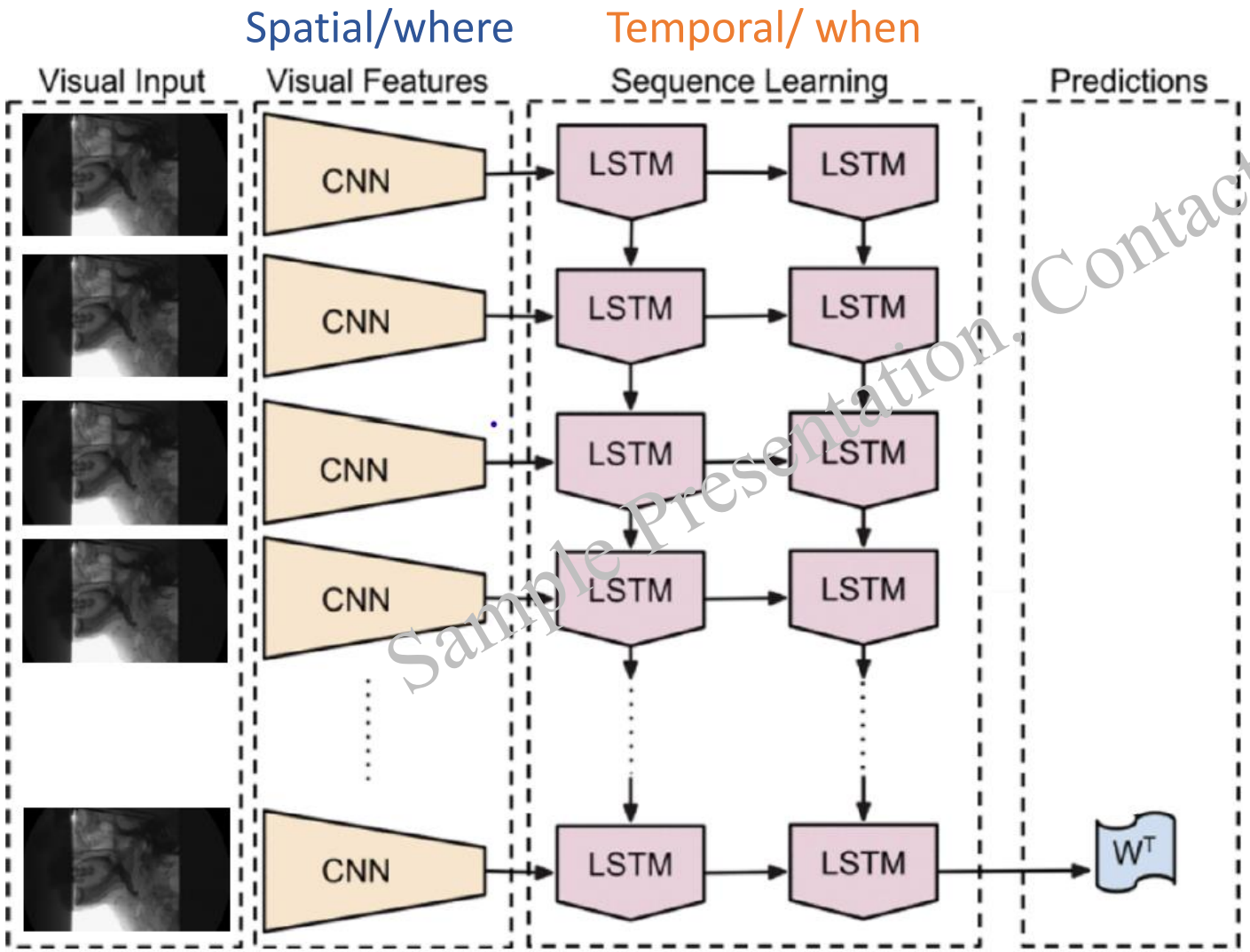
Importing Videos for ACTION tasks

	Image	**Video**
Spatial information (where?)	*Yes*	Yes
Common Tasks		*Image tasks* • *Image Classification* • *Image Detection* • *Image Generation*
Temporal information (when?)	*N.A.*	**Yes**
Unique Tasks	*Nil*	**Action tasks** • **Action recognition**

Optical flow

https://towardsdatascience.com/self-supervised-attention-mechanism-for-dense-optical-flow-estimation-b7709af48efd

Layering the inputs



https://ieeexplore.ieee.org/document/9098510

Convolution Neural Network

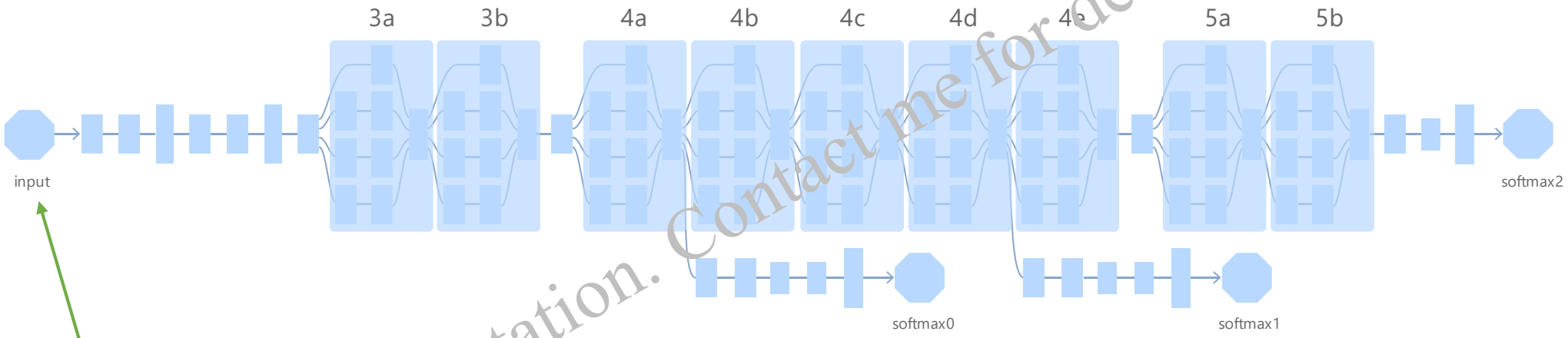


Image is fed into the model.

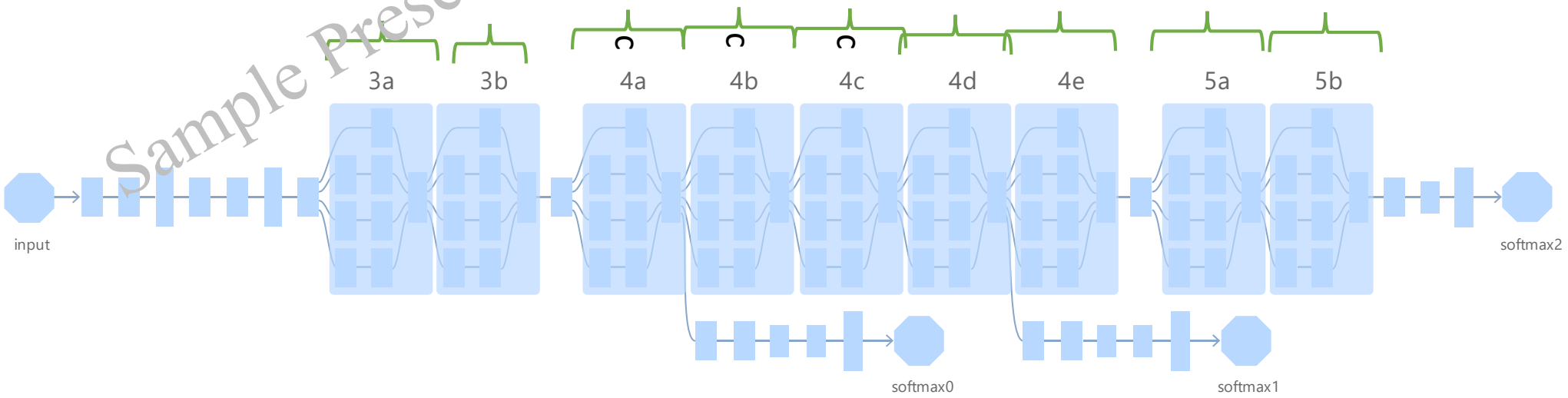
The image is read as a matrix of numbers (0-255)

https://distill.pub/2017/feature-visualization/appendix/

Convolution Neural Network

The image goes through multiple layers of mathematic transformation.

At each layer, the model learns new features of the image.

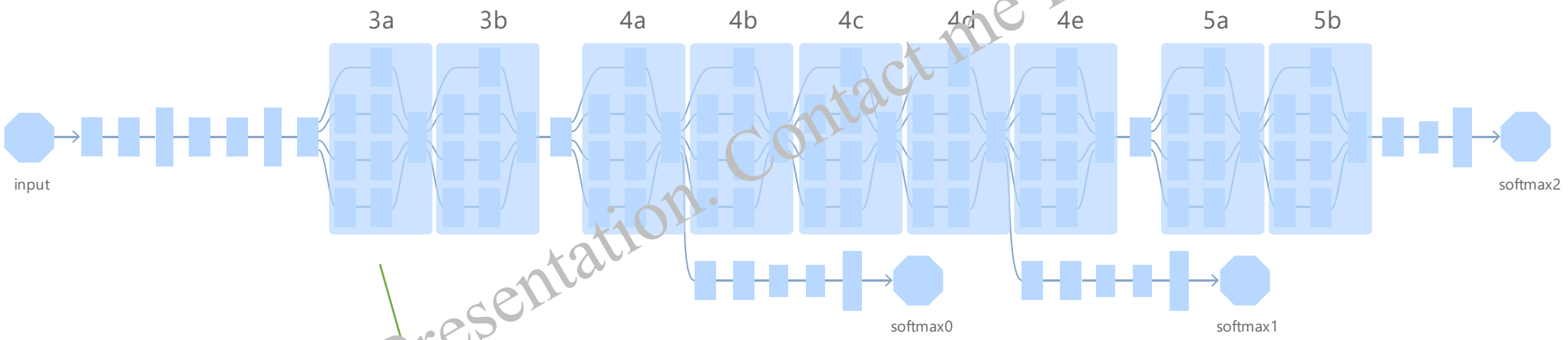


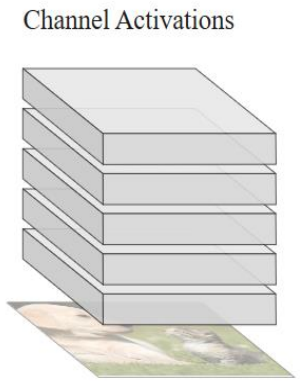
Ethical AI

- Need to address blackbox medicine
- How does CNN learn? What is CNN learning?

- Output of a layer is not a single image

- Volume of images representing all the different slices in that layer.

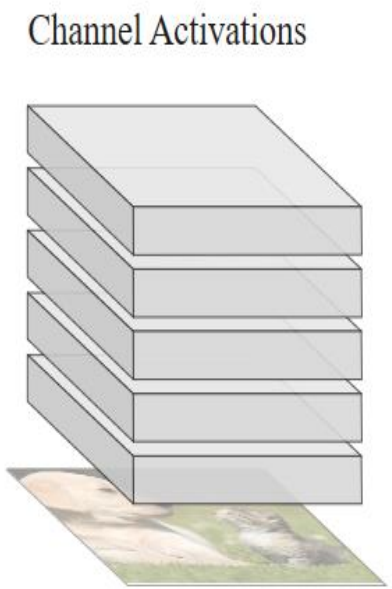


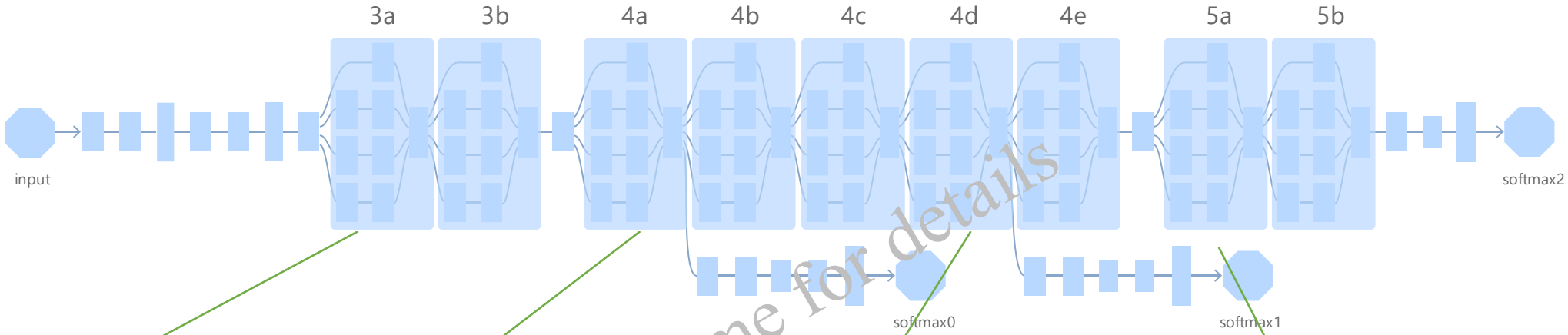


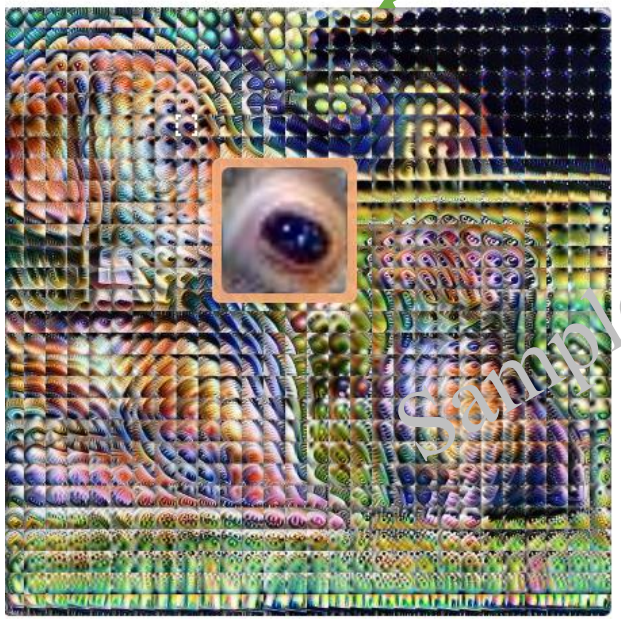
https://distill.pub/2017/feature-visualization/appendix/; https://distill.pub/2017/feature-visualization; https://distill.pub/2018/building-blocks/

Feature visualization

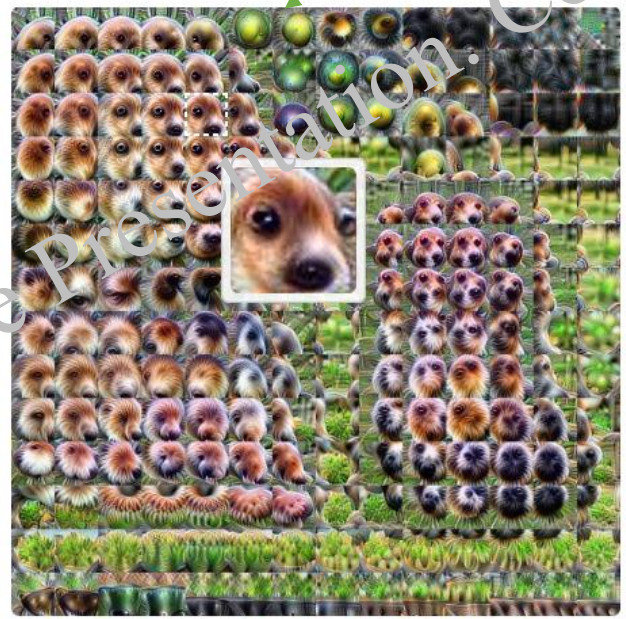
- Generate a image of what a specific slice in a specific layer is detecting.
- Creating “semantic dictionary”



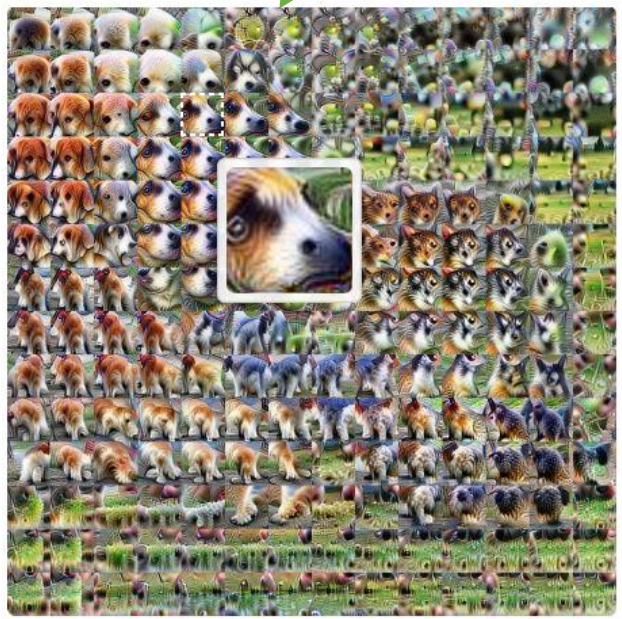




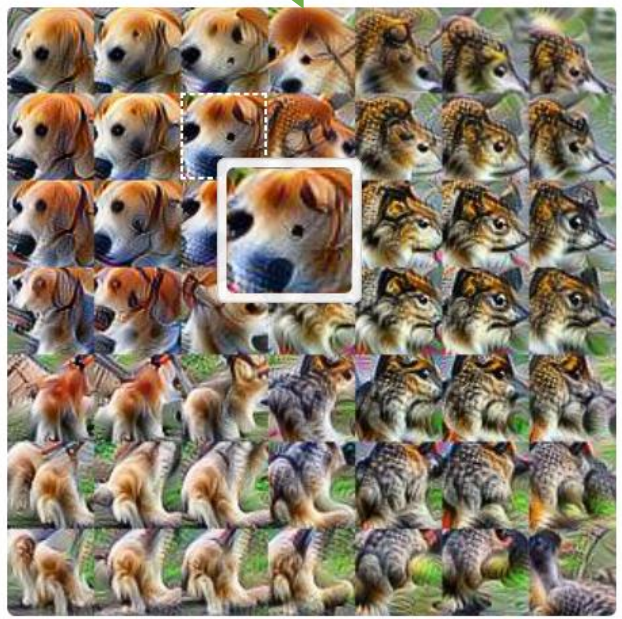
MIXED3A



MIXED4A



MIXED4D



MIXED5A

https://distill.pub/2017/feature-visualization/appendix/; https://distill.pub/2018/building-blocks/

Current clinical practice

- “Eye” power

- Subjectivity (e.g. extend of hyoid displacement)
- Time consuming

Current research practice

- Splice VFSS videos
- Annotate area of interest *(demo)*

- Reduced inter and intra-rater reliability

- Some research in using computer assistance
- Still need to annotate inputs on a calibration frame.
- Lack of scalability to operationalize for front end clinical use.

How modern CV/CNN can help

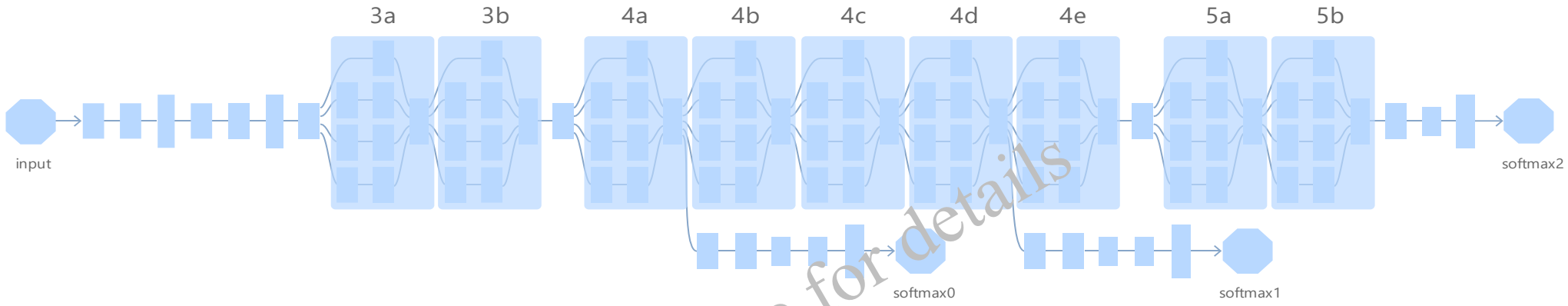
- Automaticity
- Better accuracy
- Scalability

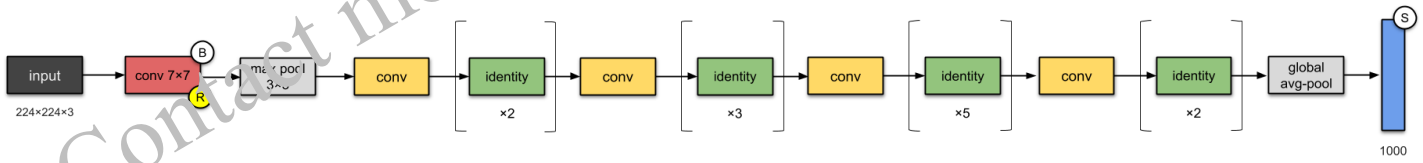
Automatic hyoid bone detection in fluoroscopic images using deep learning (2018)

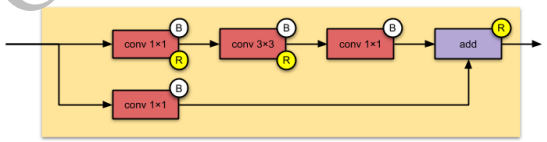
- 265 patients
- 1434 swallows
- 48,000 image frames. (70% train: 30%test)

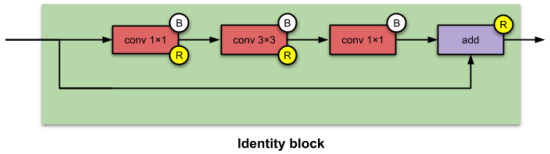
https://www.nature.com/articles/s41598-018-30182-6

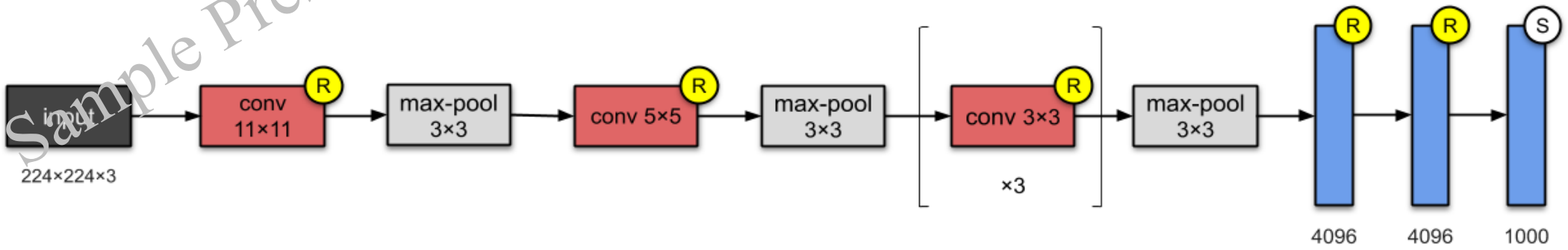
Different CNN architecture







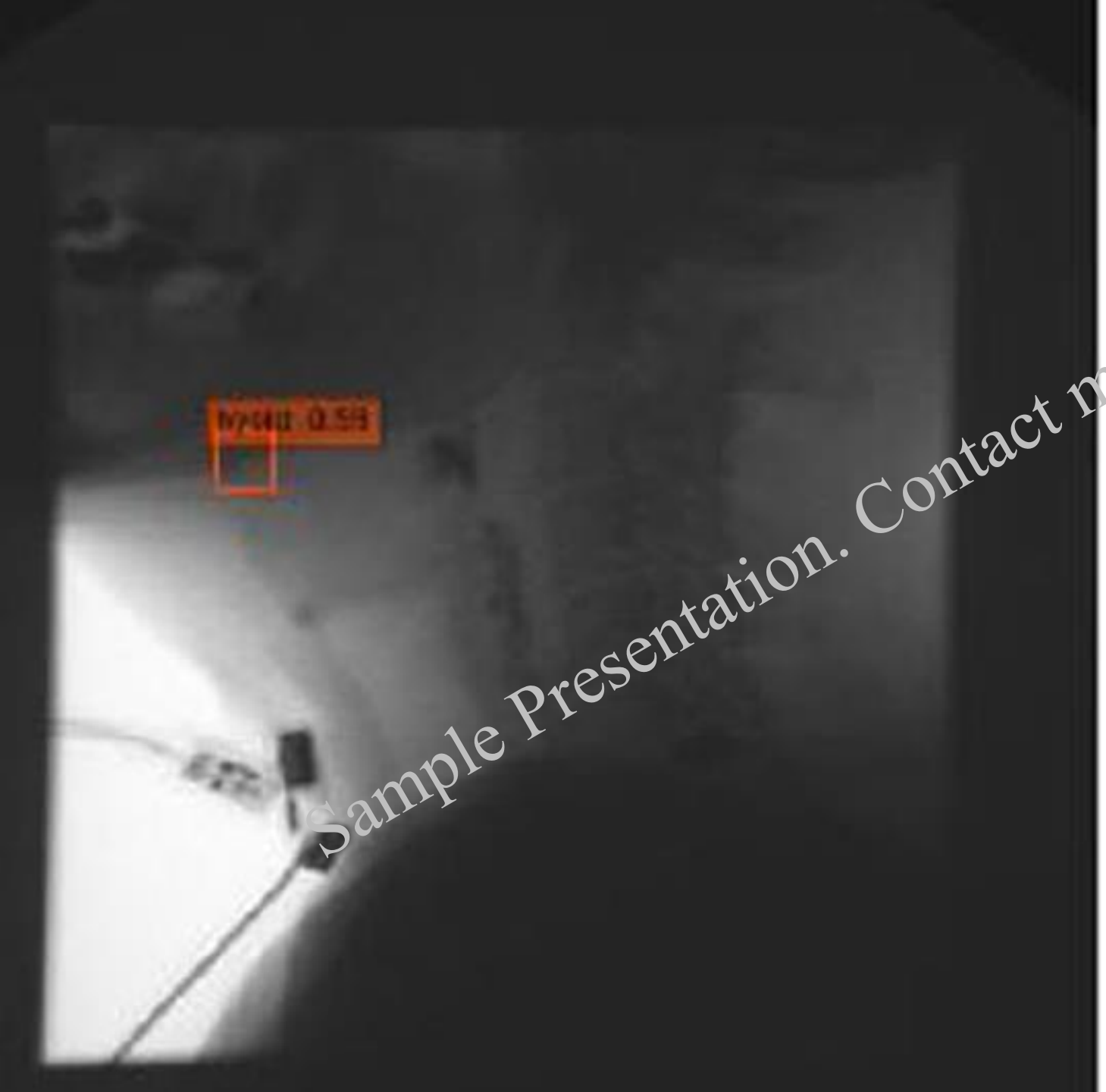




https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d

Clinical findings

Model	Mean average precision
YOLOv2	33.10%
Faster-RCNN + ZF	69.01%
SSD300-VGG	84.37%
SSD300-ResNet	81.92%
SSD500-VGG	89.14%
SSD500-ResNet	89.03%



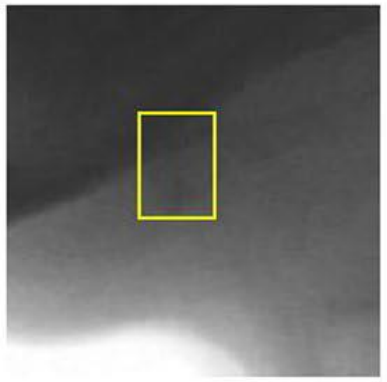
Potential application

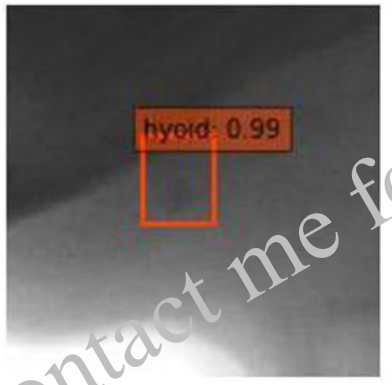
- automatic segmentation of hyoid bone areas
- determine hyoid displacement is disordered
- Is reduced airway protection or reduced UES 2’ hyo-laryngeal excursion?

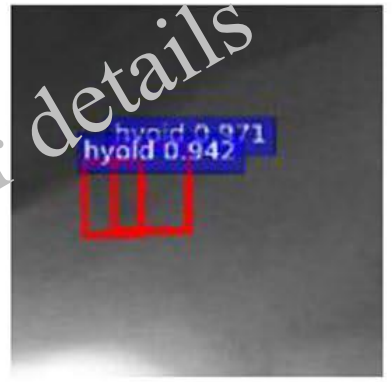
ML findings

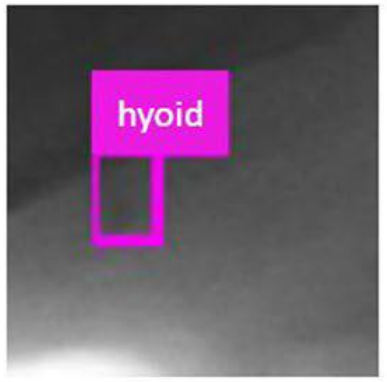
- Radiation dosage
- Model's ability to detect small objects vs computational time

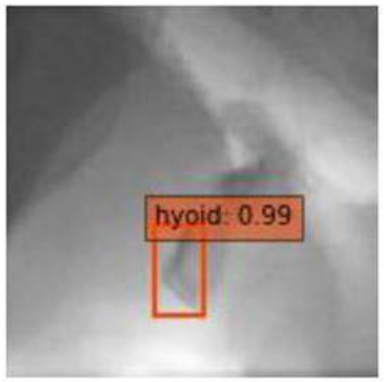
- Unexpected detections
  - Colour boxes are detected hyoid bones

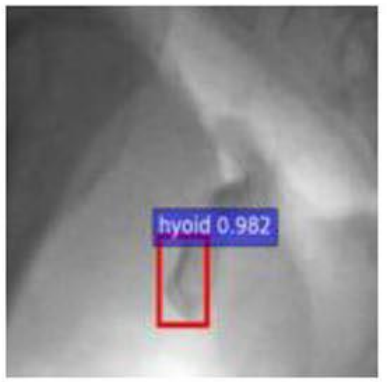


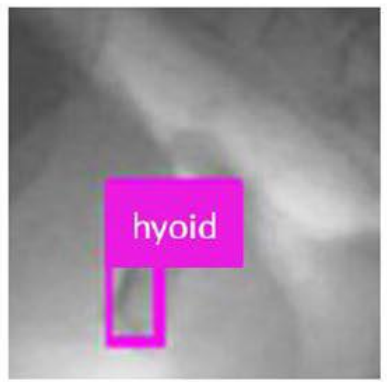












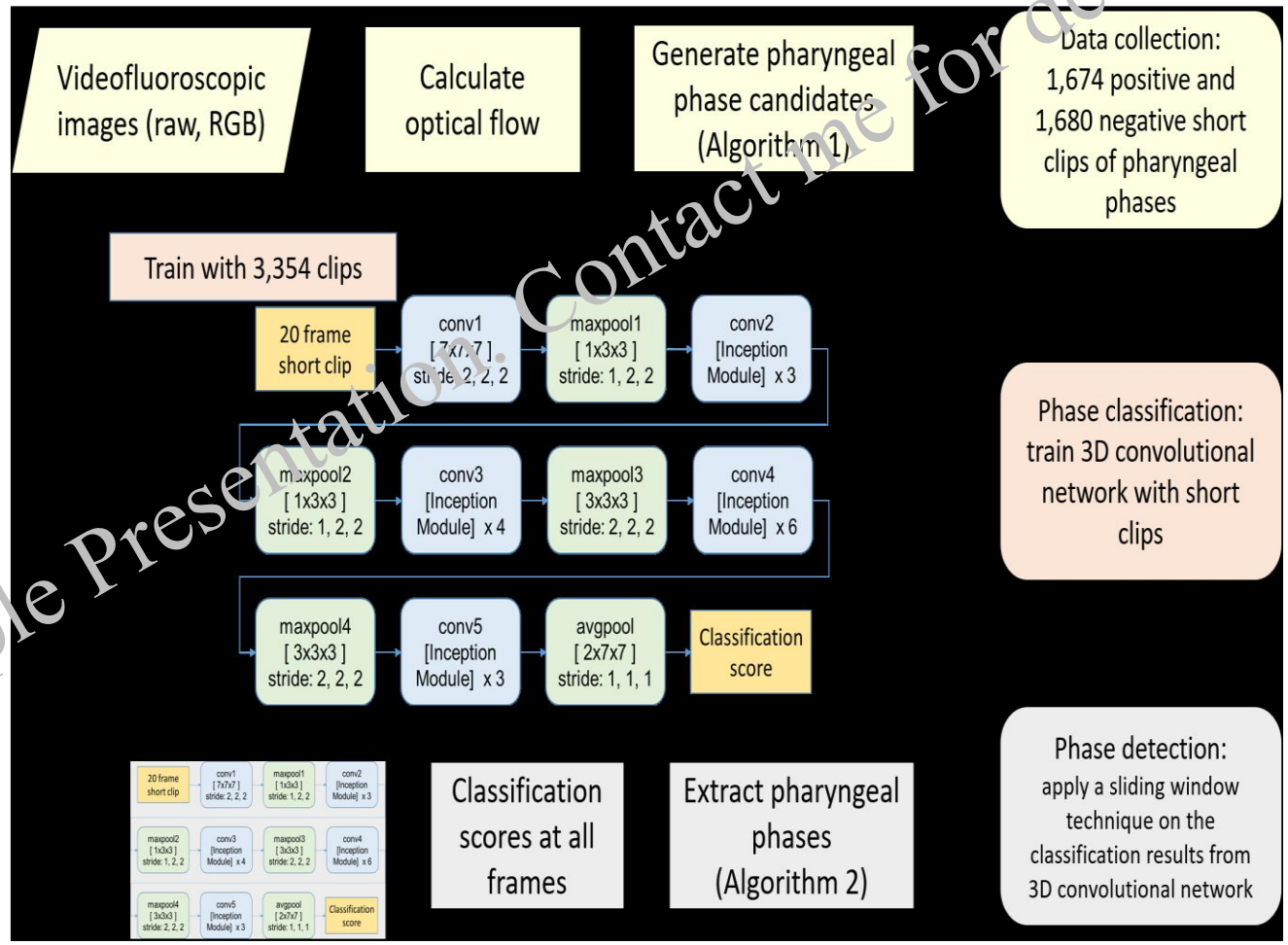
Automatic Detection of the Pharyngeal Phase in VFSS Using Efficient Data Collection and 3D Convolutional Networks (2019)

- Pharyngeal phase <1s
  - Finding a needle in a haystack

- identifies the pharyngeal phase in VFSS
- without the need for spatial or temporal annotations.

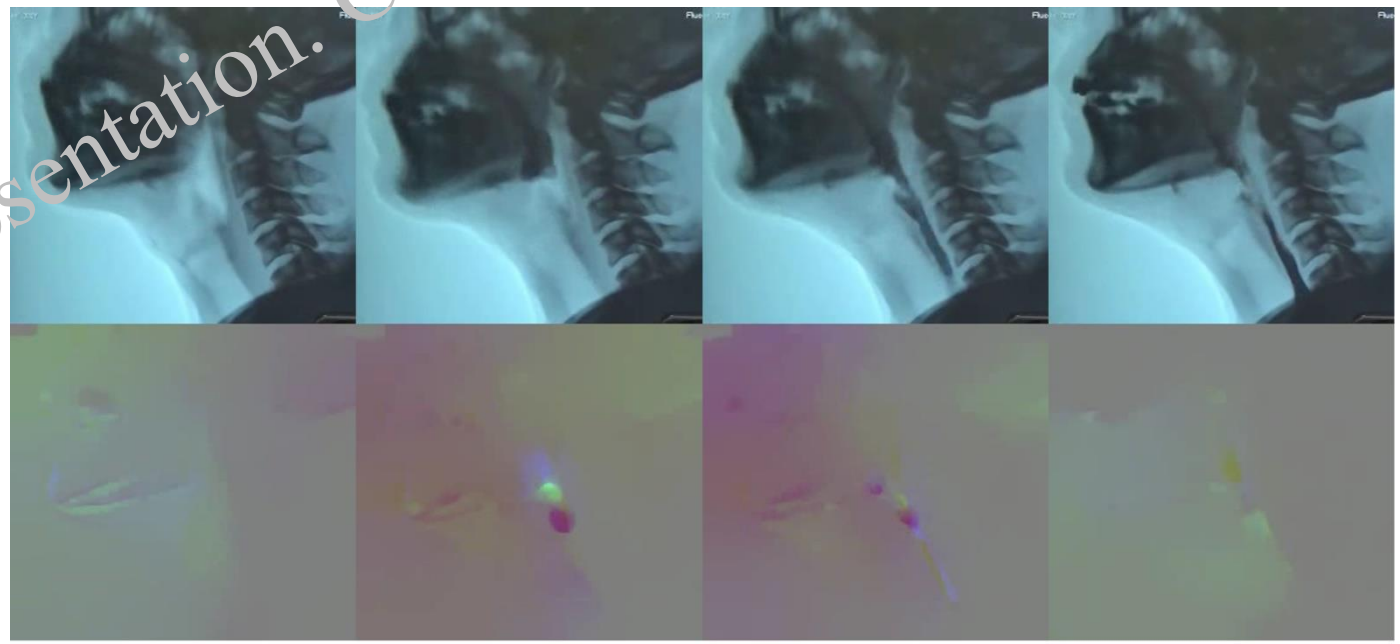
https://www.mdpi.com/1424-8220/19/18/3873

3 stage process



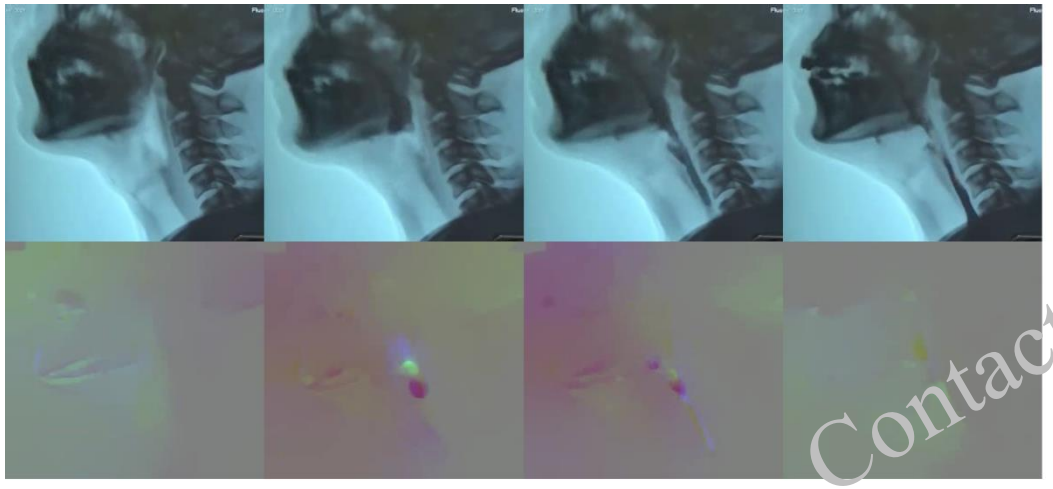
Phase 1

- Models used can be computational taxing
- Screen VFSS video for pharyngeal frames
- Optical flow to detect vertical placement of pixels

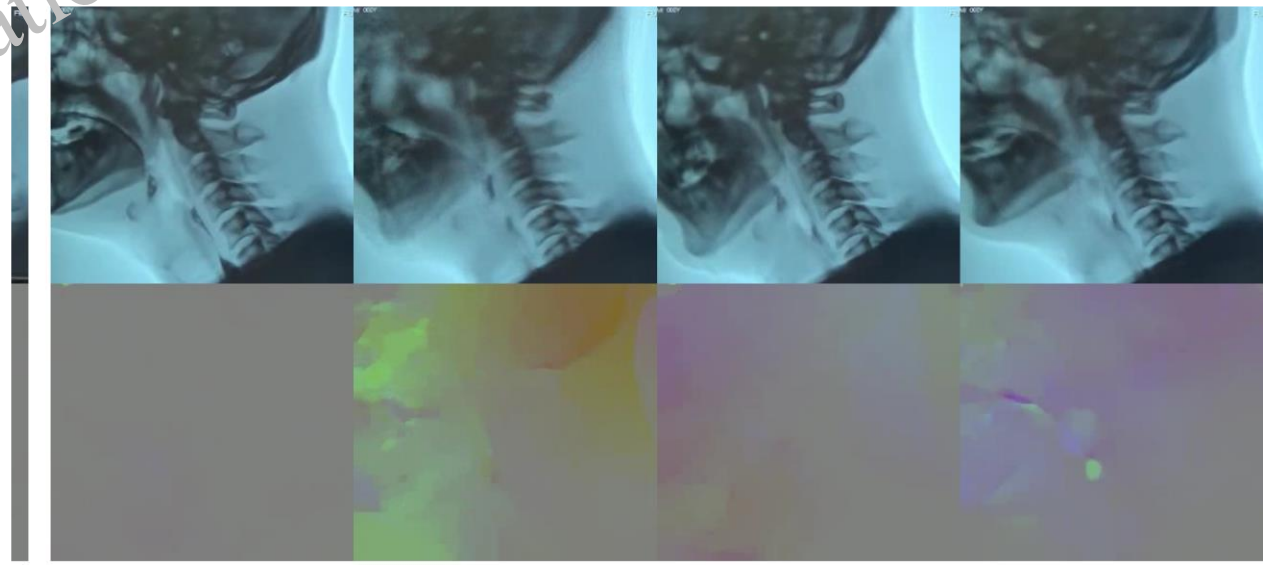


(a) Pharyngeal phase

False positive

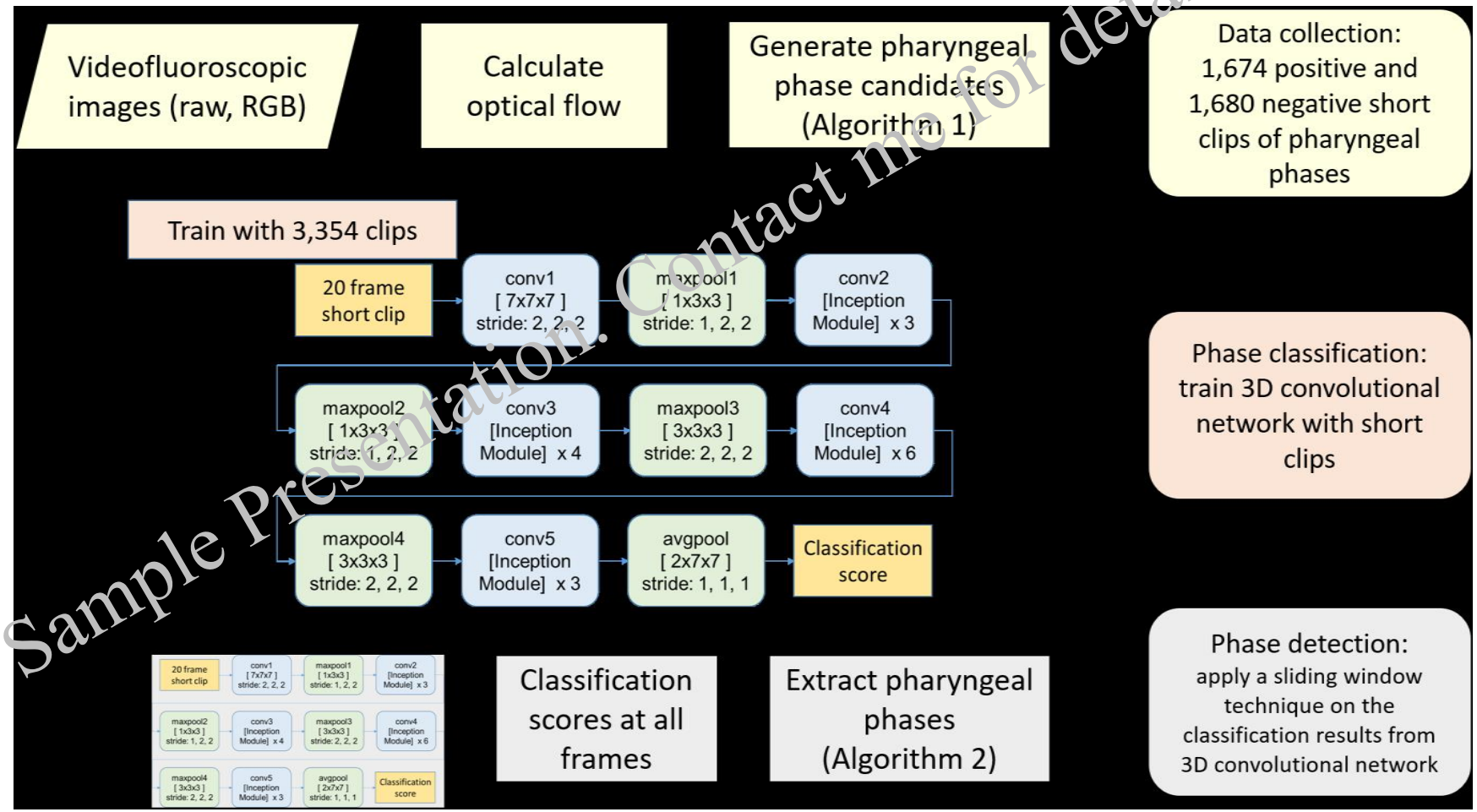


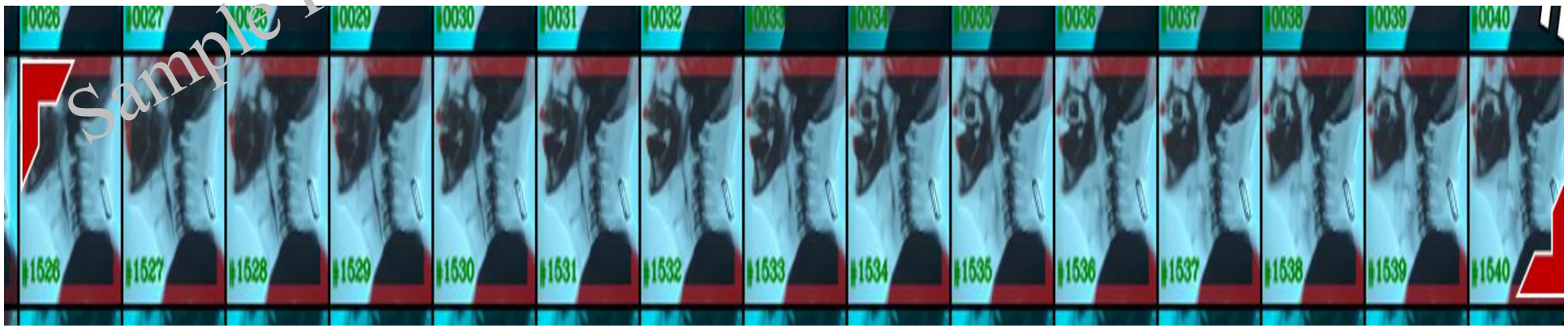
(a) Pharyngeal phase



(b) Head shaking

Phase 2-3





- Detection Score (F1)
  - 45.73 to 84.25%

- Detection Time Error
  - 1.42s to 3.82s

Others

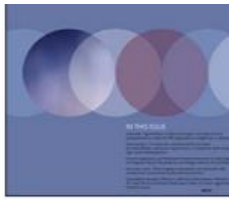
- Deep learning based application for videofluoroscopic swallowing study (VFSS): A pilot study (2019)
  - PAS

- A deep learning approach to VFSS exam classification (2020)
  - Abnormal swallow

https://www.jns-journal.com/article/S0022-510X(19)30829-9/pdf; https://ieeexplore.ieee.org/document/9098510

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad



Review

The future of radiology augmented with Artificial Intelligence: A strategy for success

Charlene Liew

Department of Radiology, Changi General Hospital, 2 Simei Street 3, 529889, Singapore

ARTICLE INFO

Keywords:
Artificial Intelligence
Deep learning
Informatics
Business
Strategy

ABSTRACT

The rapid development of Artificial Intelligence/deep learning technology and its implementation into routin clinical imaging will cause a major transformation to the practice of radiology. Strategic positioning will ensur the successful transition of radiologists into their new roles as augmented clinicians. This paper describes a overall vision on how to achieve a smooth transition through the practice of augmented radiology wher radiologists-in-the-loop ensure the safe implementation of Artificial Intelligence systems.

https://www.sciencedirect.com/science/article/abs/pii/S0720048X18301116

The future of radiology augmented with Artificial Intelligence: A strategy for success (2018)

1. Current state and the need for a strategy
2. General use cases, potential impact and implementation strategy
3. Impact upon cost leadership, differentiation and focus
4. Defining roles, technical considerations and requirements for implementation
5. Organizational aspects of implementation
6. Special considerations, job displacement and risk mitigation
7. Safety, privacy, moral and ethical concerns