U-Net & BERT

Olaf Ronneberger & Jacob Devlin

2026.01.05

BrainLAB Journal Club
Department of Applied Artificial Intelligence
Jeong SangYeop

Overview

01 Author & Journal

02 Challenges

03 U-Net

04 BERT

01 Author & Journal (U-Net)

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Homepage of Olaf Ronneberger

apl. Prof. Dr. Olaf Ronneberger

Google DeepMind
London, UK
Twitter: @ORonneberger

and

Albert-
Institu
Lehrst
George
D-7911

Email:

U-net: Convolutional networks for biomedical image segmentation O Ronneberger, P Fischer, T Brox International Conference on Medical image computing and computer-assisted ...	126184	2015
Highly accurate protein structure prediction with AlphaFold J Jumper, R Evans, A Pritzel, T Green, M Figurnov, O Ronneberger, ... nature 596 (7873), 583-589	44506	2021
Accurate structure prediction of biomolecular interactions with AlphaFold 3 J Abramson, J Adler, J Dunger, R Evans, T Green, A Pritzel, ... Nature 630 (8016), 493-500	10679	2024
3D U-Net: learning dense volumetric segmentation from sparse annotation Ö Çiçek, A Abdulkadir, SS Lienkamp, T Brox, O Ronneberger International conference on medical image computing and computer-assisted ...	10293	2016
Protein complex prediction with AlphaFold-Multimer R Evans, M O'Neill, A Pritzel, N Antropova, A Senior, T Green, A Žídek, ... biorxiv, 2021.10. 04.463034	3691	2021
Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context G Team, P Georgiev, VI Lei, R Burnell, L Bai, A Gulati, G Tanzer, ... arXiv preprint arXiv:2403.05530	3307	2024

01 Author & Journal (BERT)

BrAIn Lab.
Brain and Artificial Intelligence Lab.

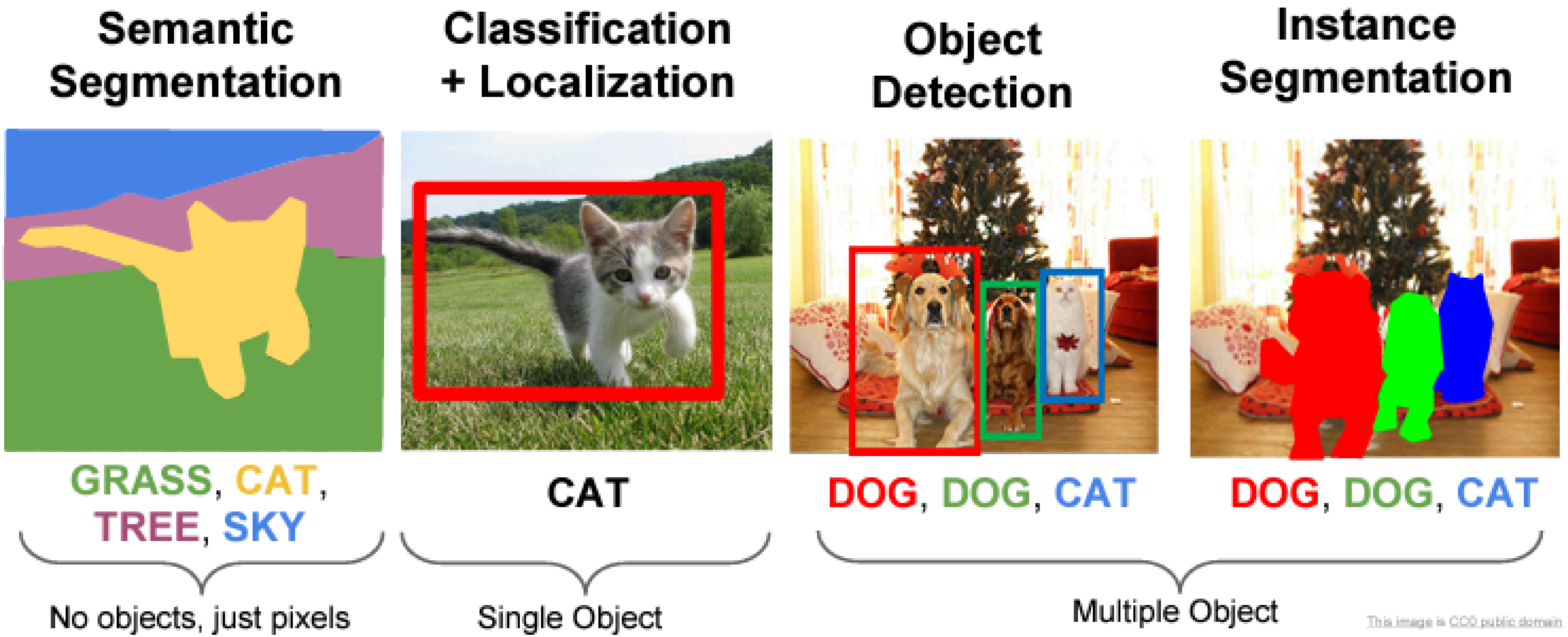
Jacob Devlin
Software Engineer at Google
미국 워싱턴 레드먼드 · 연락처

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
Jacob Devlin | Ming-Wei Chang | Kenton Lee | Kristina Toutanova · Computer Science ·
North American Chapter of the Association for... · 2019
TLDR A new language representation model, BERT, designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers, which can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks. Expand
107,456 · 21,679 · PDF · ACL · Save · Alert · Cite

Natural Questions: A Benchmark for Question Answering Research
T. Kwiatkowski | J. Palomaki | +15 authors | Slav Petrov · Computer Science ·
Transactions of the Association for Computational... · 1 August 2019
TLDR The Natural Questions corpus, a question answering data set, is presented, introducing robust metrics for the purposes of evaluating question answering systems; demonstrating high human upper bounds on these metrics; and establishing baseline results using competitive methods drawn from related literature. Expand
4,049 · 511 · PDF · ACL · Save · Alert · Cite

Scaling Instruction-Finetuned Language Models
Hyung Won Chung | Le Hou | +29 authors | Jason Wei · Computer Science ·
Journal of machine learning research · 20 October 2022
TLDR It is found that instruction finetuning with the above aspects dramatically improves performance on a variety of model classes (PaLM, T5, U-PaLM), prompting setups, and evaluation benchmarks (MMLU, BBH, TyDiQA, MGSM, open-ended generation). Expand
3,773 · 411 · PDF · arXiv · Save · Alert · Cite

02 Several Challenges for Computer Vision



"Pixel-wise classification"

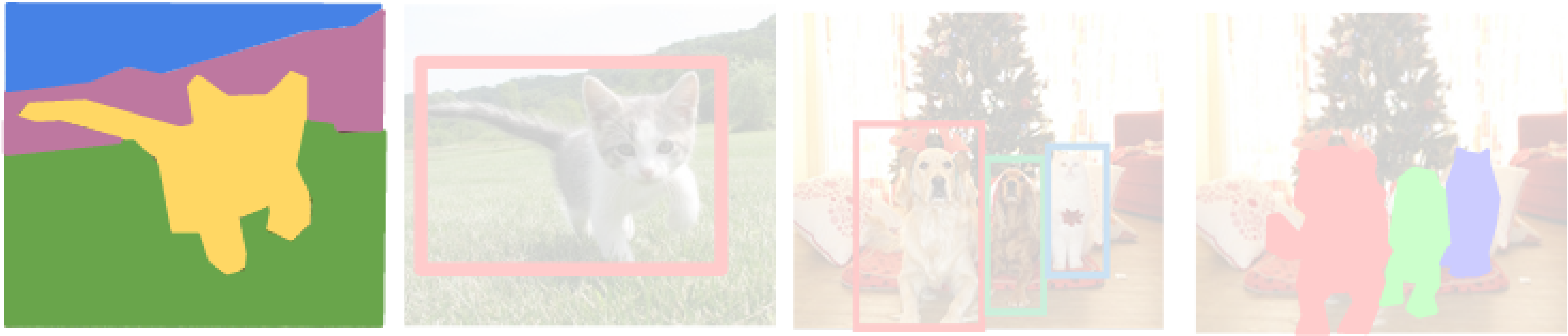
"Bounding box localization + classification"

"Object Detection + Semantic Segmentation"

02 Several Challenges for Computer Vision

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Semantic Segmentation | Classification + Localization | Object Detection | Instance Segmentation



GRASS, CAT, TREE, SKY | CAT | DOG, DOG, CAT | DOG, DOG, CAT

No objects, just pixels | Single Object | Multiple Object

This image is CC0 public domain

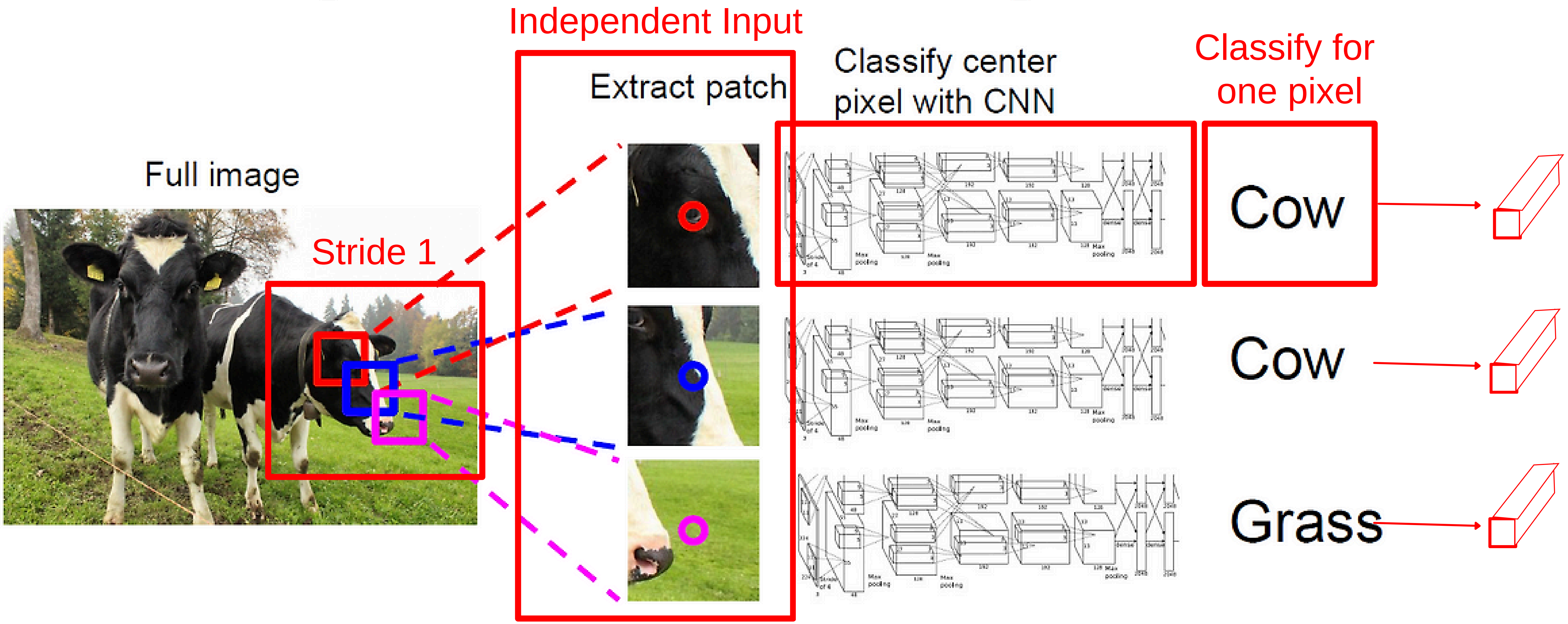
“Pixel-wise classification”

“Bounding box localization + classification”

“Object Detection + Semantic Segmentation”

Brain and Artificial Intelligence Lab.

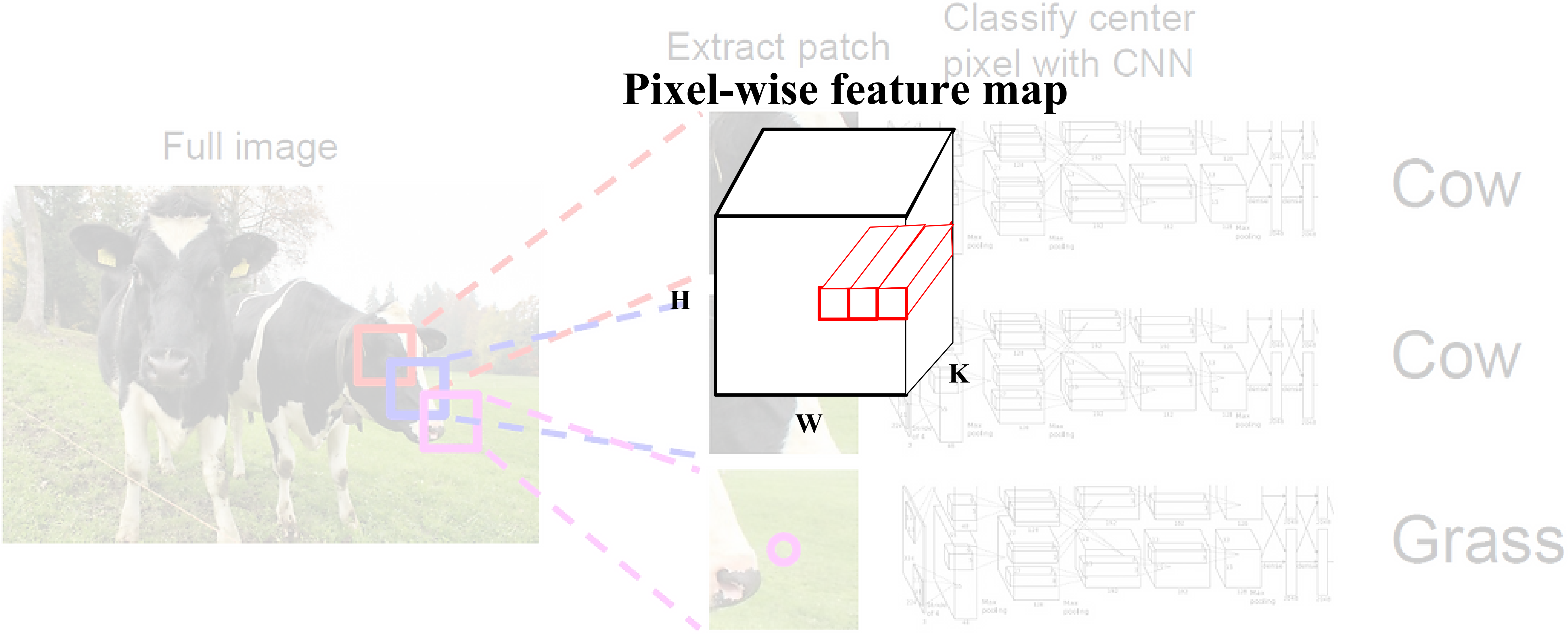
Semantic Segmentation Idea: Sliding Window



Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

02 Segmentation Challenge – Sliding window

Semantic Segmentation Idea: Sliding Window



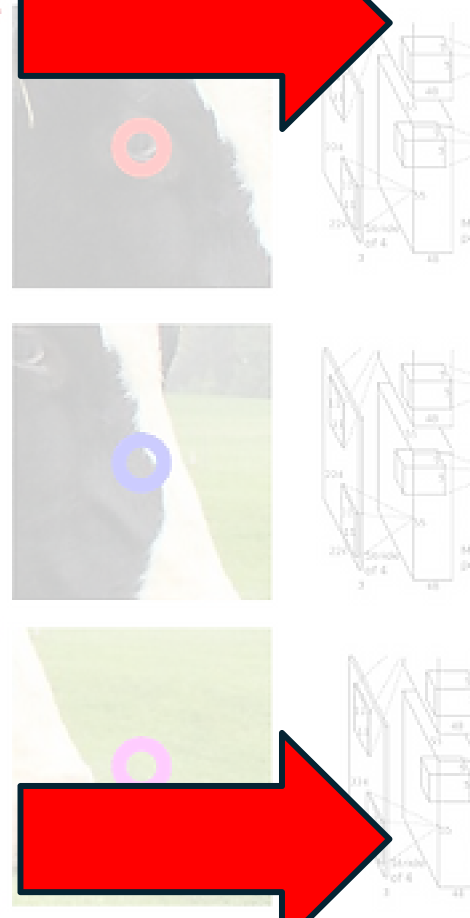
Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

02 Segmentation Challenge – Sliding window

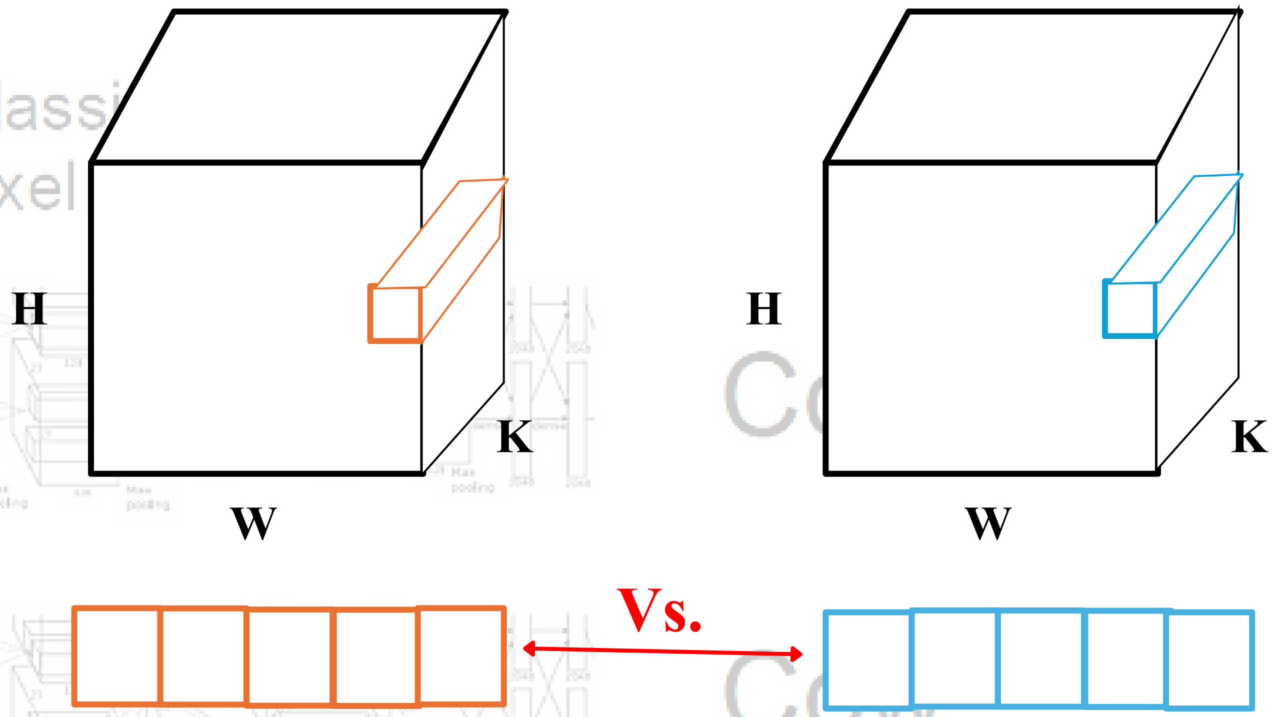
Semantic Segmentation Idea: Sliding Window

Softmax

- **x: specific pixel coordinate**
- **(x): Activation value in x (in k channel)**
- **whole Class**

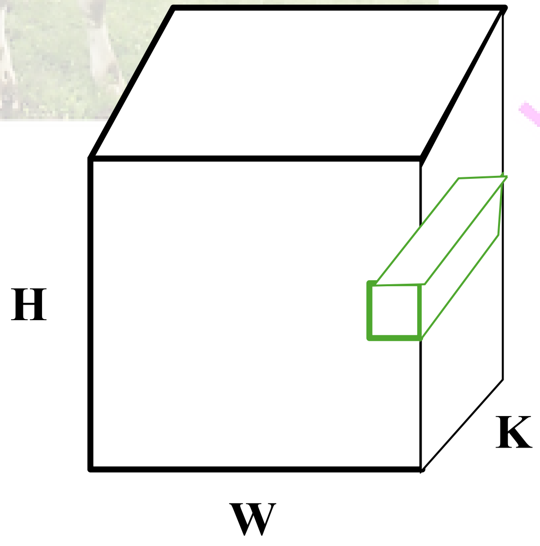


Probability map Vs. Ground Truth



Loss function

Cross Entropy



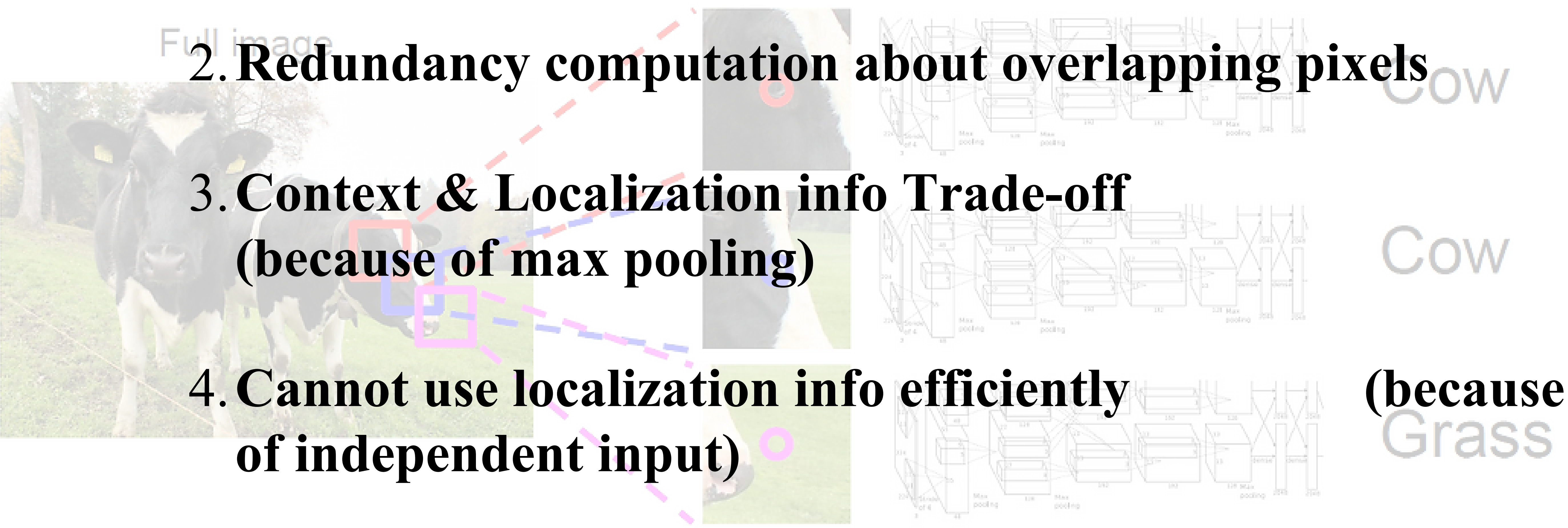
"Total Loss"

Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Drawbacks

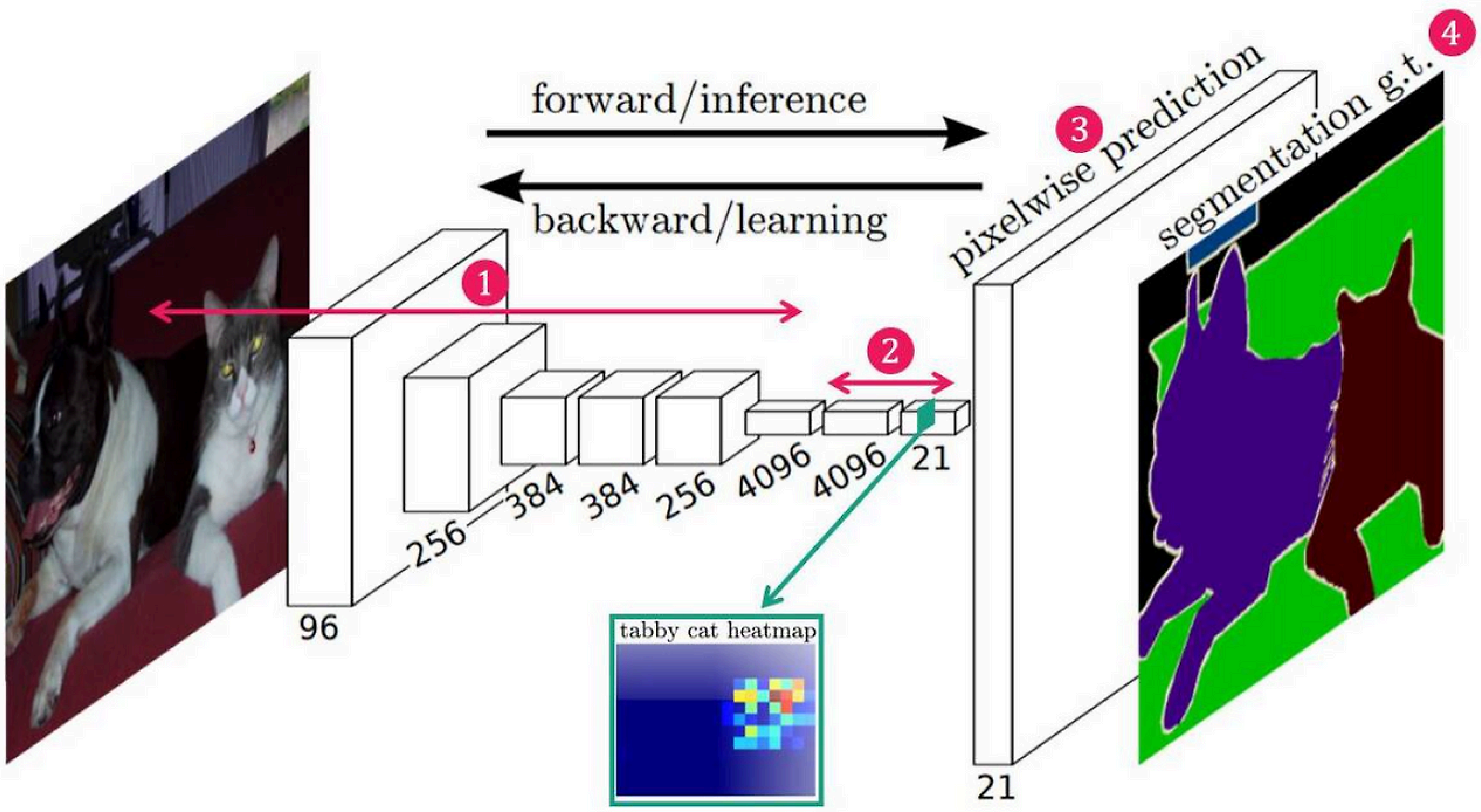
1. Shift 1 pixel each time for Probability map(model result)
2. Redundancy computation about overlapping pixels
3. Context & Localization info Trade-off (because of max pooling)
4. Cannot use localization info efficiently (because of independent input)



Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

02 Segmentation Challenge - FCN

BrAIn Lab.
Brain and Artificial Intelligence Lab.

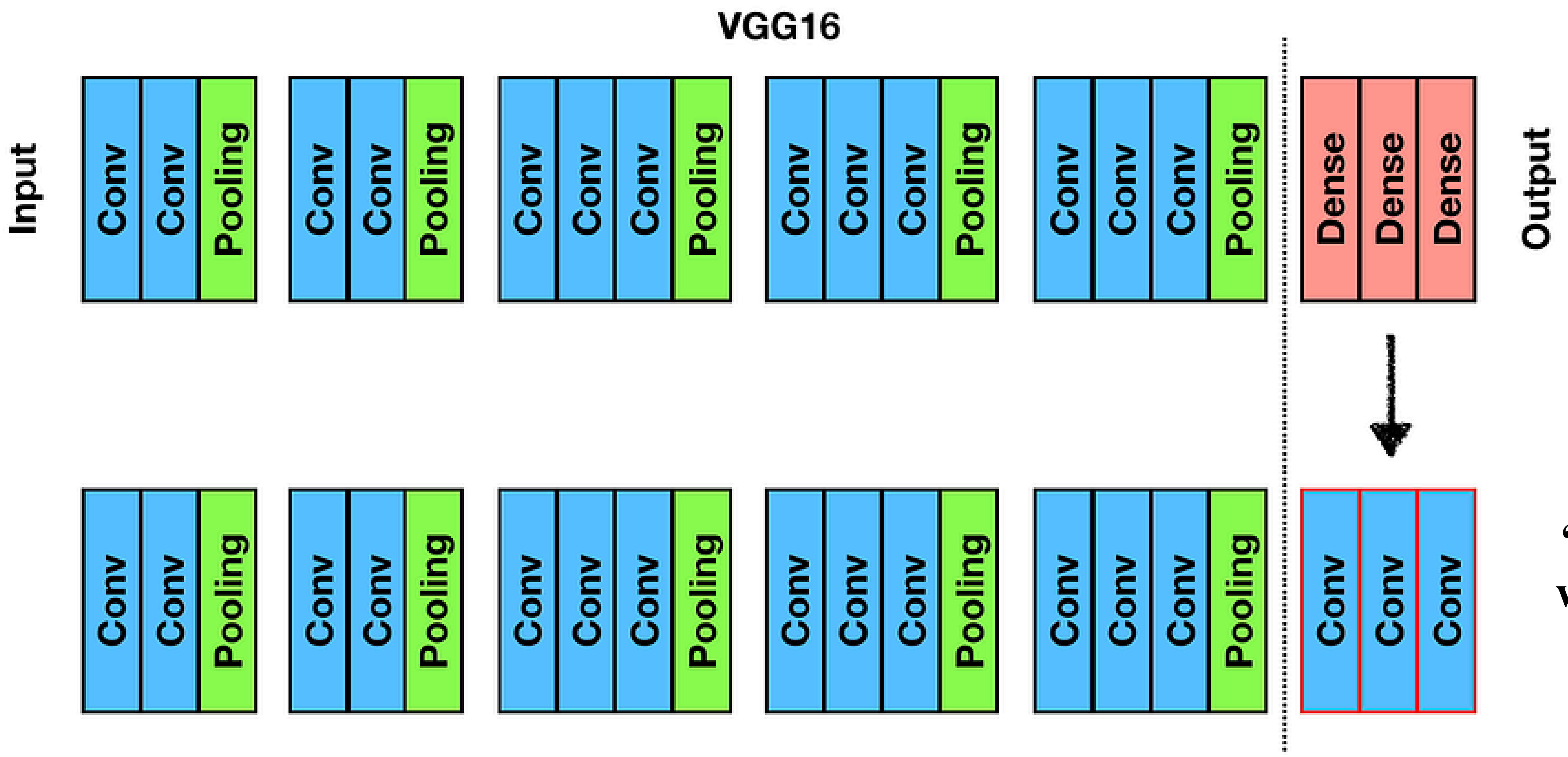


Attributes

1. Contracting path & Expansive Path
2. Whole Image goes into the input
3. No Fully connected layer (replaced with 1*1 convolution layers)
4. Typically implemented by VGG-Net
5. Using “Skip Connection + Addition”

02 Segmentation Challenge - FCN

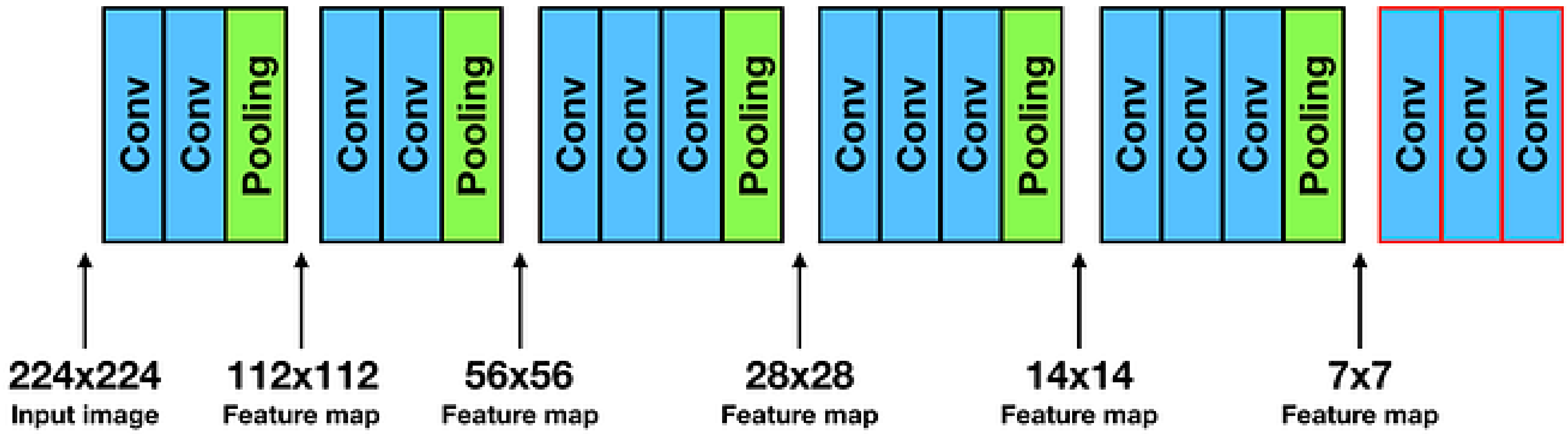
BrAIn Lab.
Brain and Artificial Intelligence Lab.



replaced FC layers with 7*7 Convolution layer, 1*1 Convolution layer, 1*1 Convolution layer!

"Skip connection and Addition" will be mentioned after few slides

02 Segmentation Challenge - FCN



1. **7*7 convolution layer: resolution info fusion (fc layer role)**

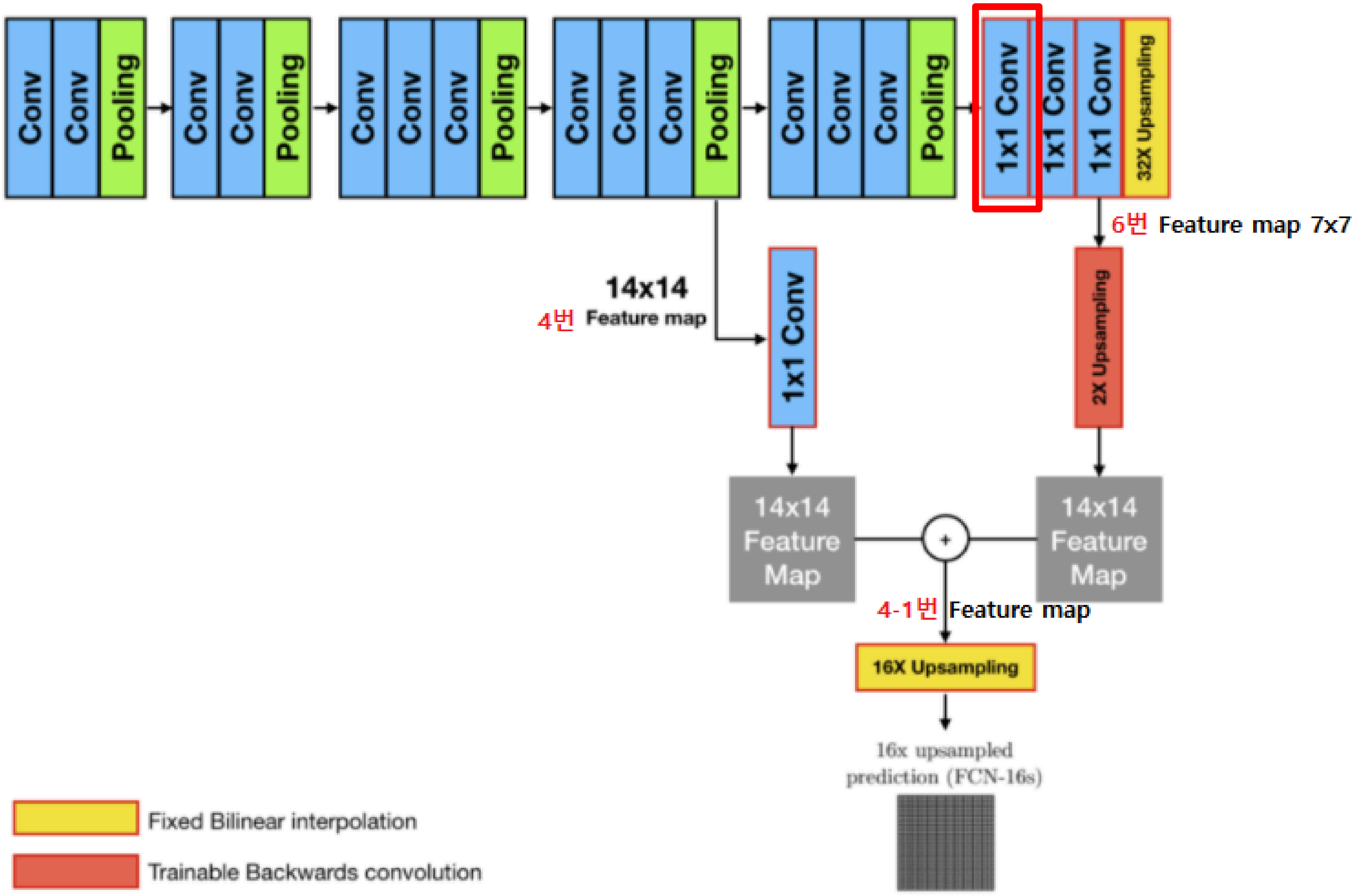
2. **1*1 convolution layer: channel info fusion (fc layer role)**

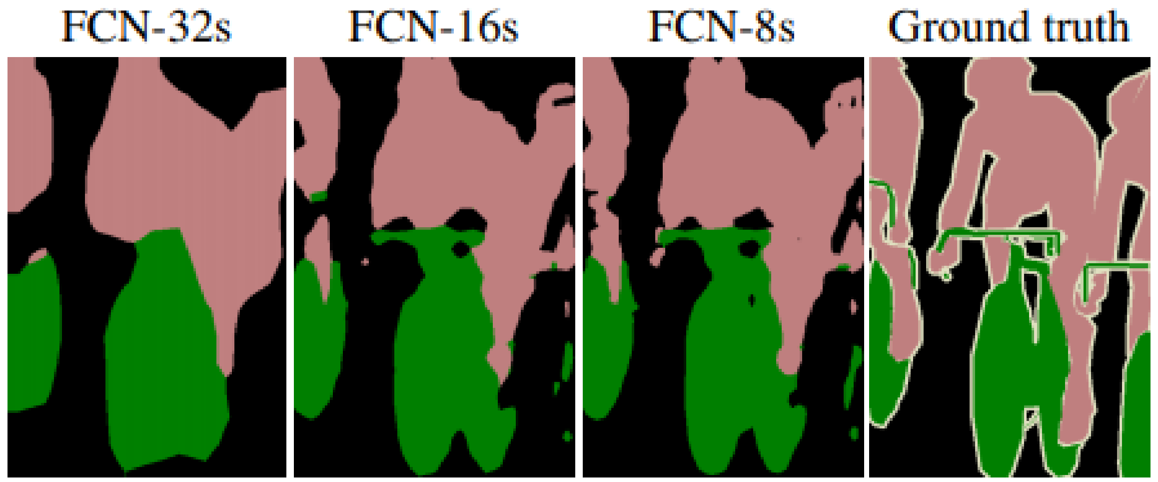
3. **1*1 convolution layer: change the number of channels (same with class numbers)**

02 Segmentation Challenge - FCN

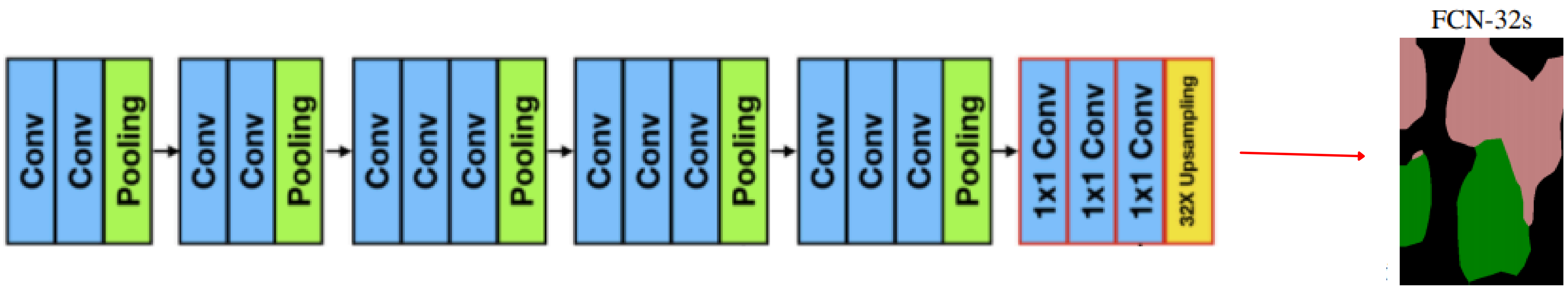
BrAIn Lab.
Brain and Artificial Intelligence Lab.

?: It is 7*7



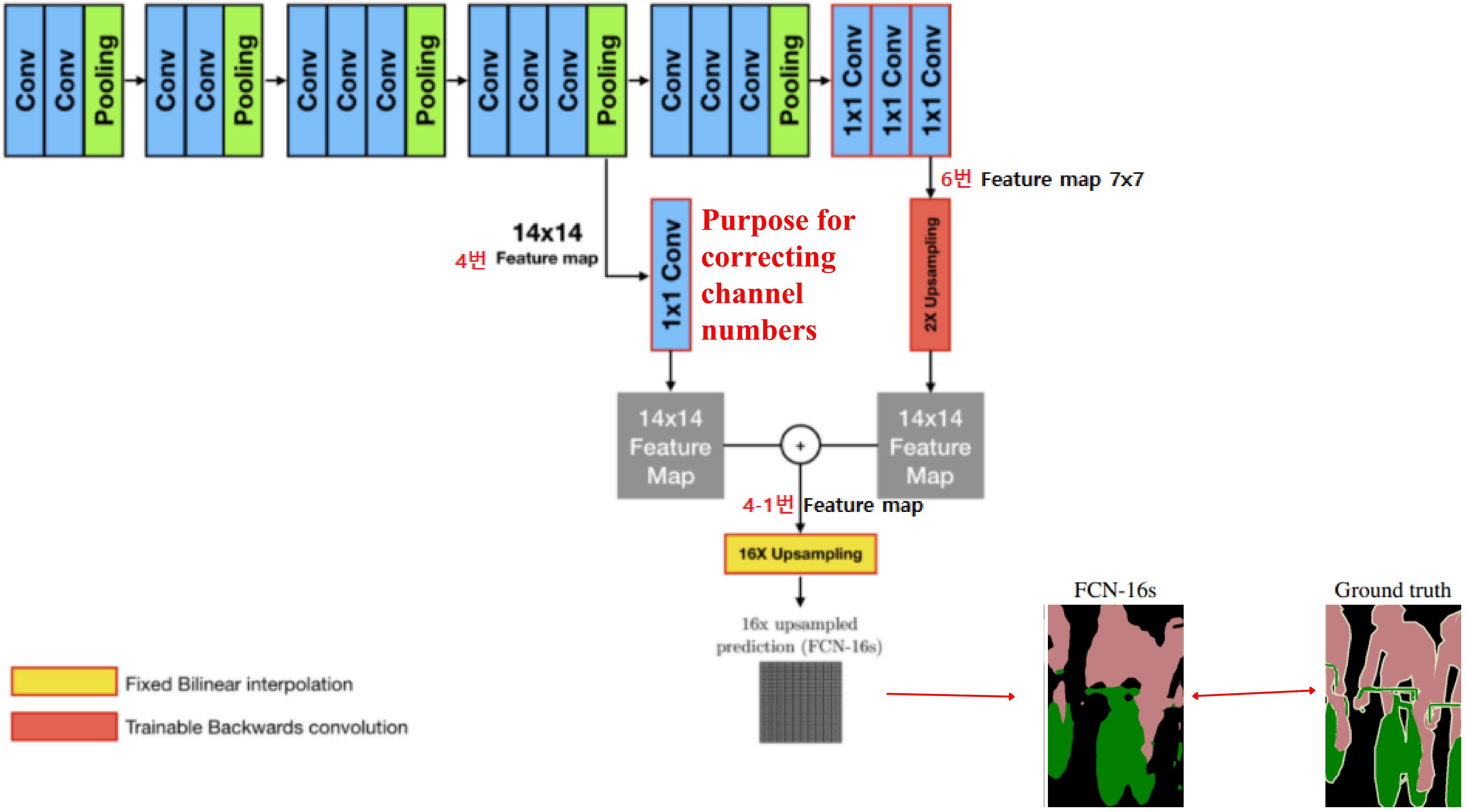


02 Segmentation Challenge - FCN32s



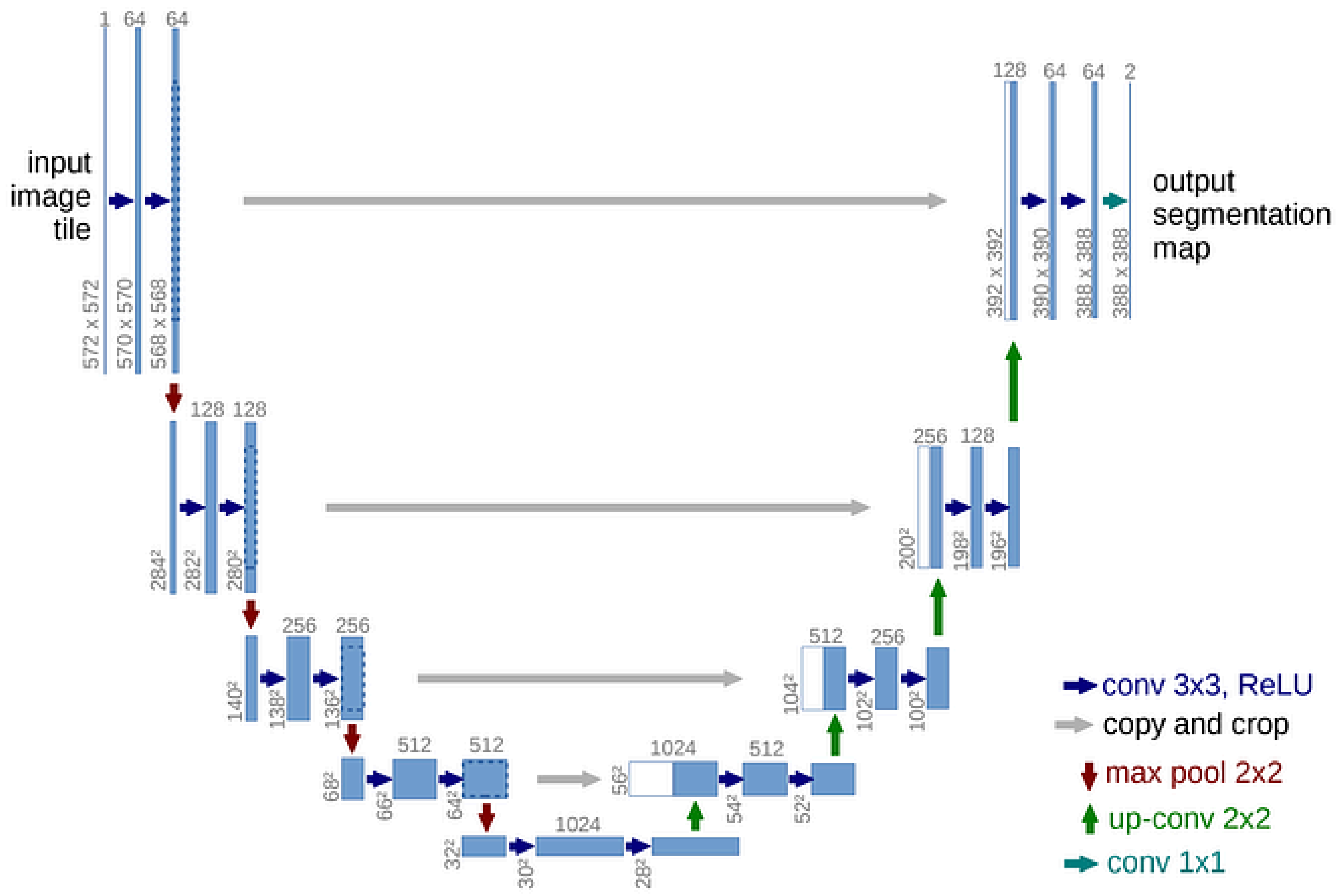
- **32X Upsampling (Transposed convolution layer)**
- **This version don’t use skip connection and addition**
- **Image quality is poor**

02 Segmentation Challenge - FCN16s



03 U-Net vs. FCN

BrAIn Lab.
Brain and Artificial Intelligence Lab.



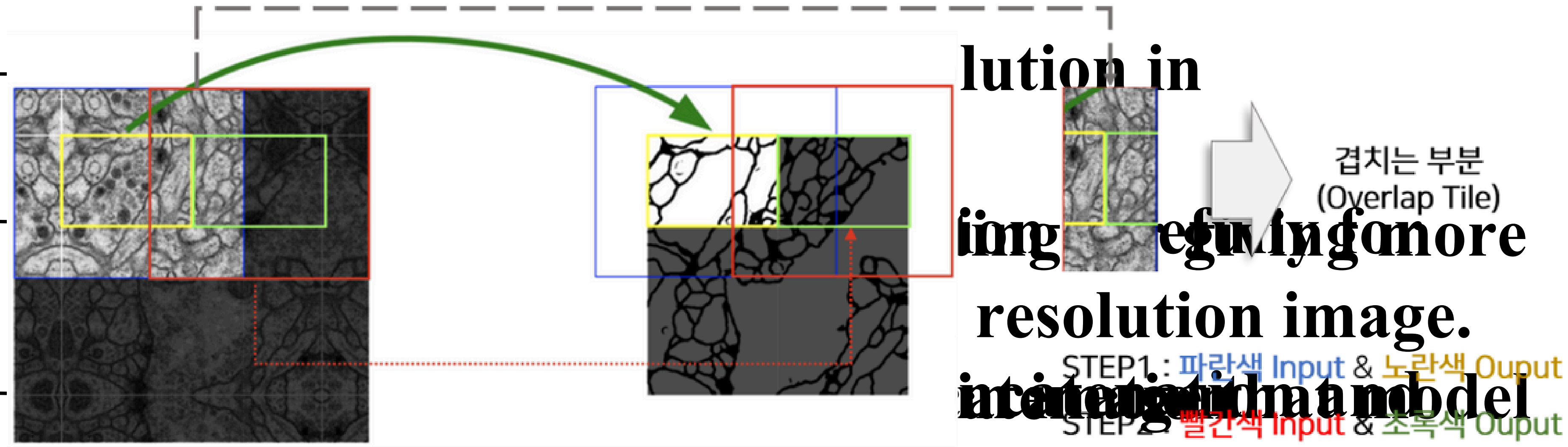
Attributes

1. **No padding**
2. **Expansive Path is more complicated**
3. **Using Concatenation (not Addition)**
4. **Using tile image (similar with patch)**

03 U-Net vs. FCN

BrAIn Lab.
Brain and Artificial Intelligence Lab.

1. No padding
2. Expansive Path is more complicated
3. Using Concatenation (not Addition)
4. Using tile image (similar with patch)

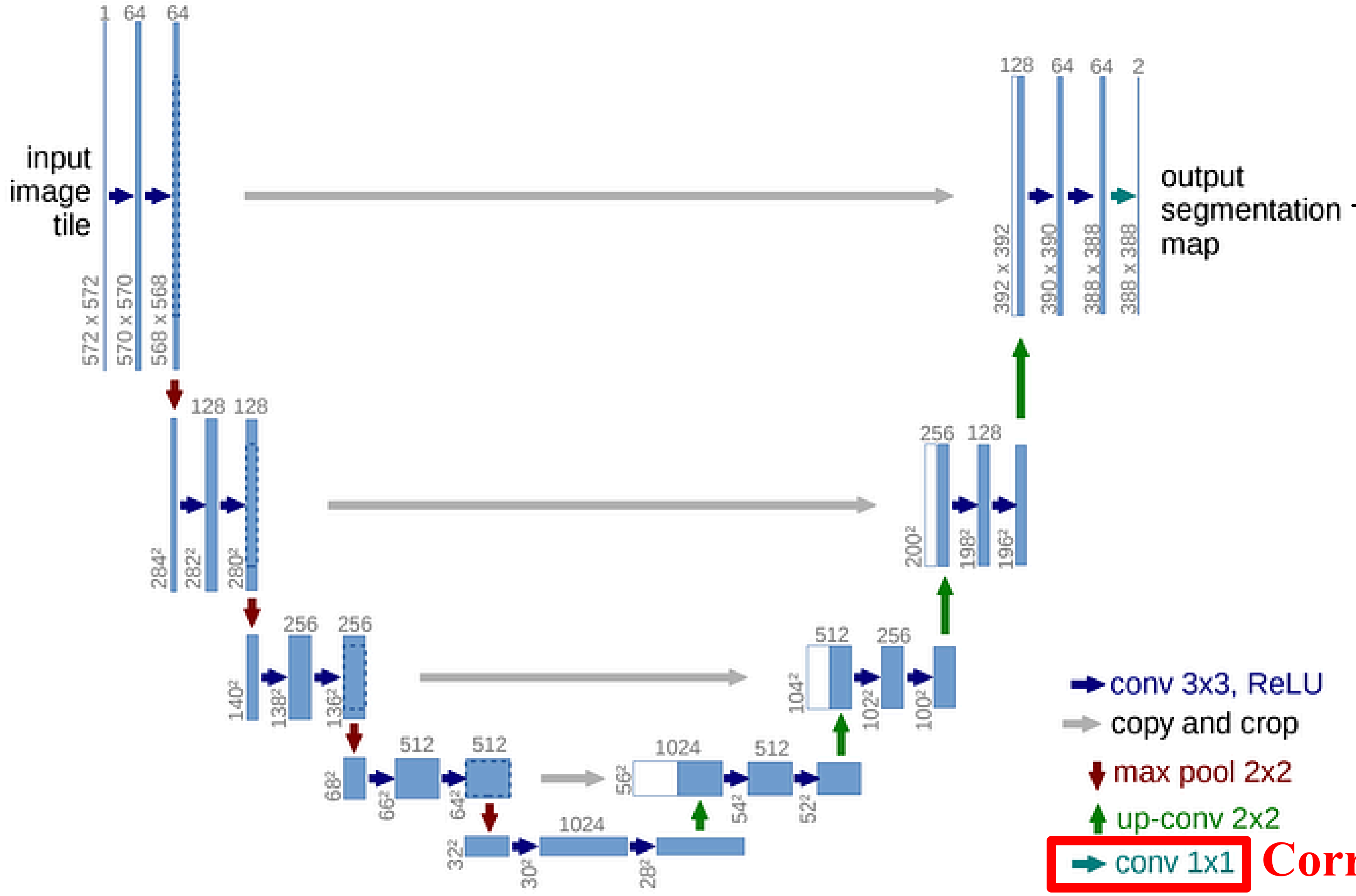


lution in

resolution image.

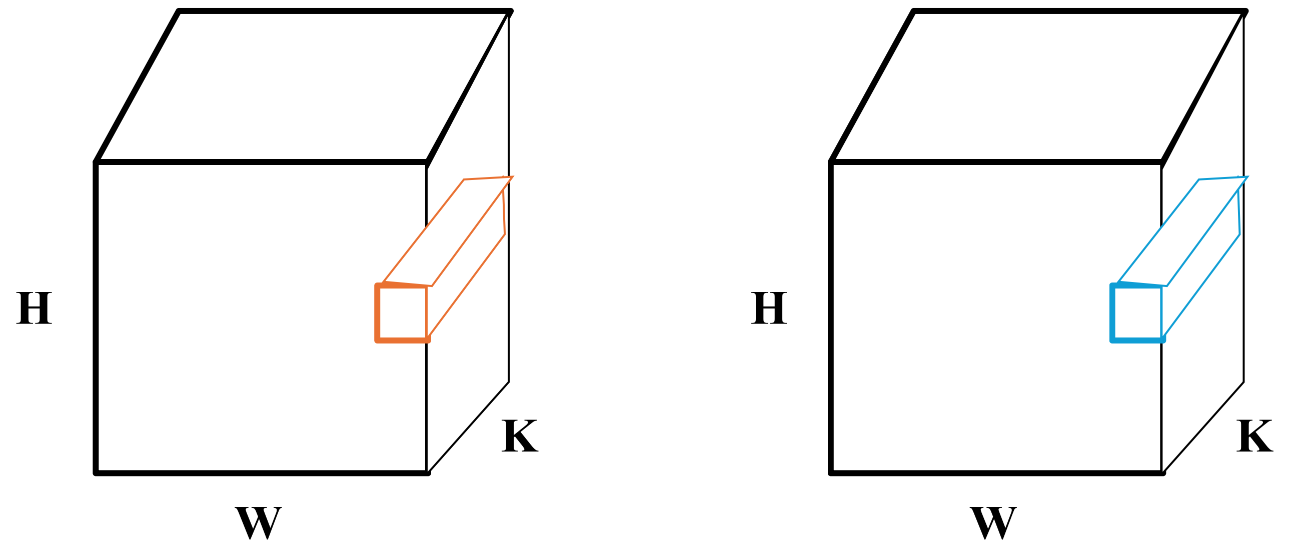
- the concatenated feature map and
- each other, the final segamantation map(final result) would be seamless

03 U-Net. Training



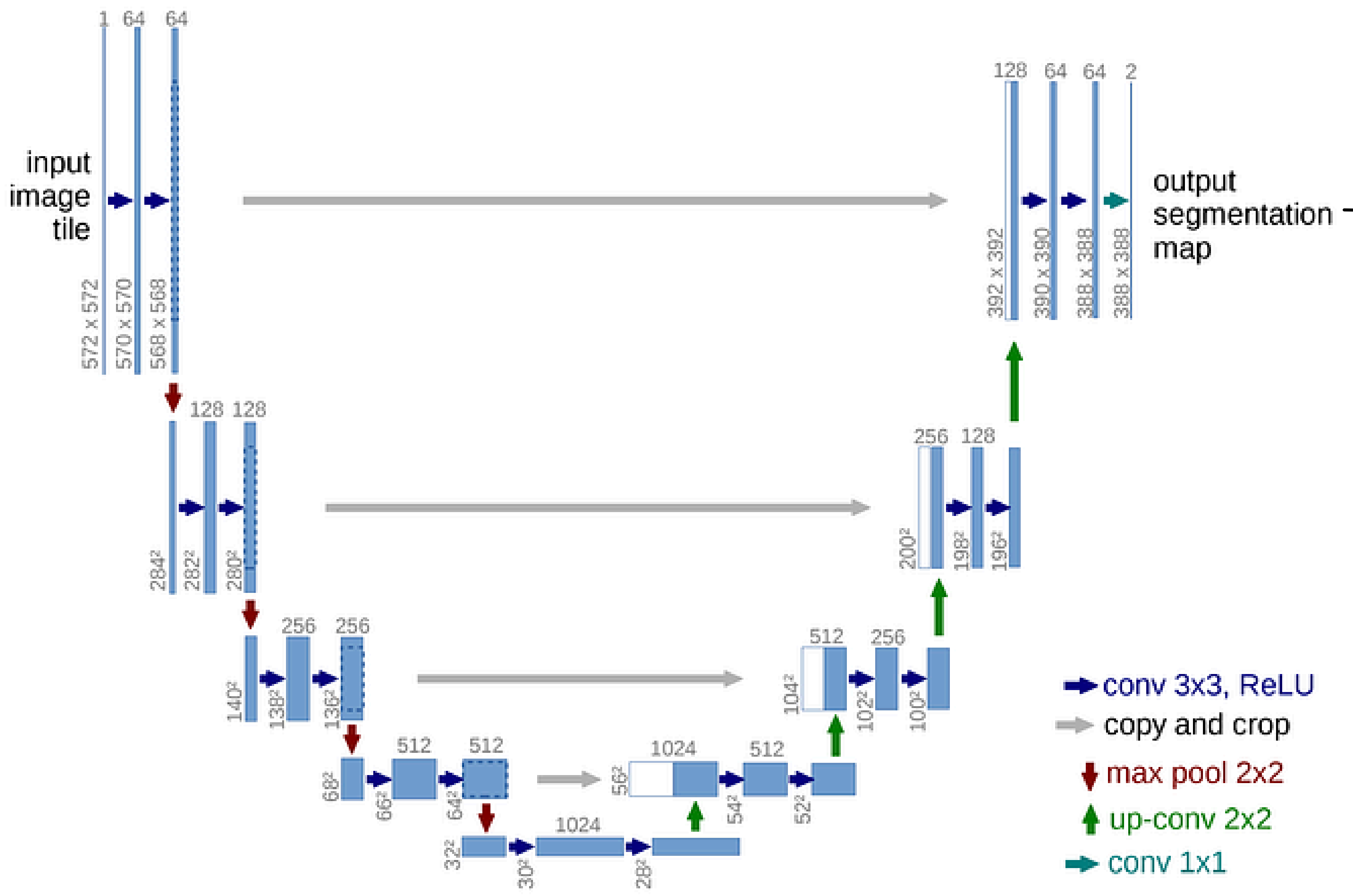
Softmax

Probability map **Vs.** **Ground Truth**

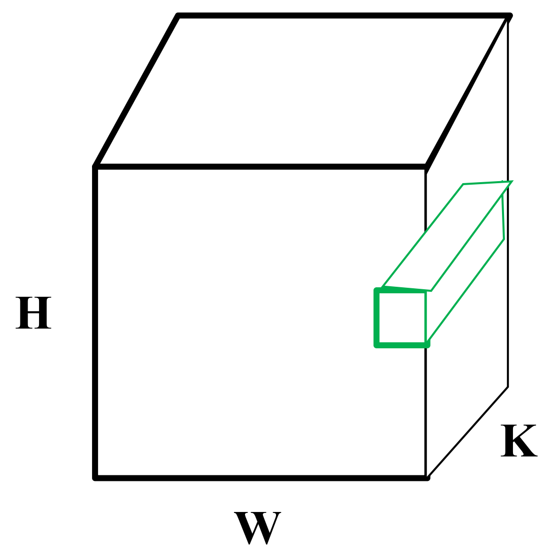


Correcting Channel numbers
= Class numbers

03 U-Net. Training

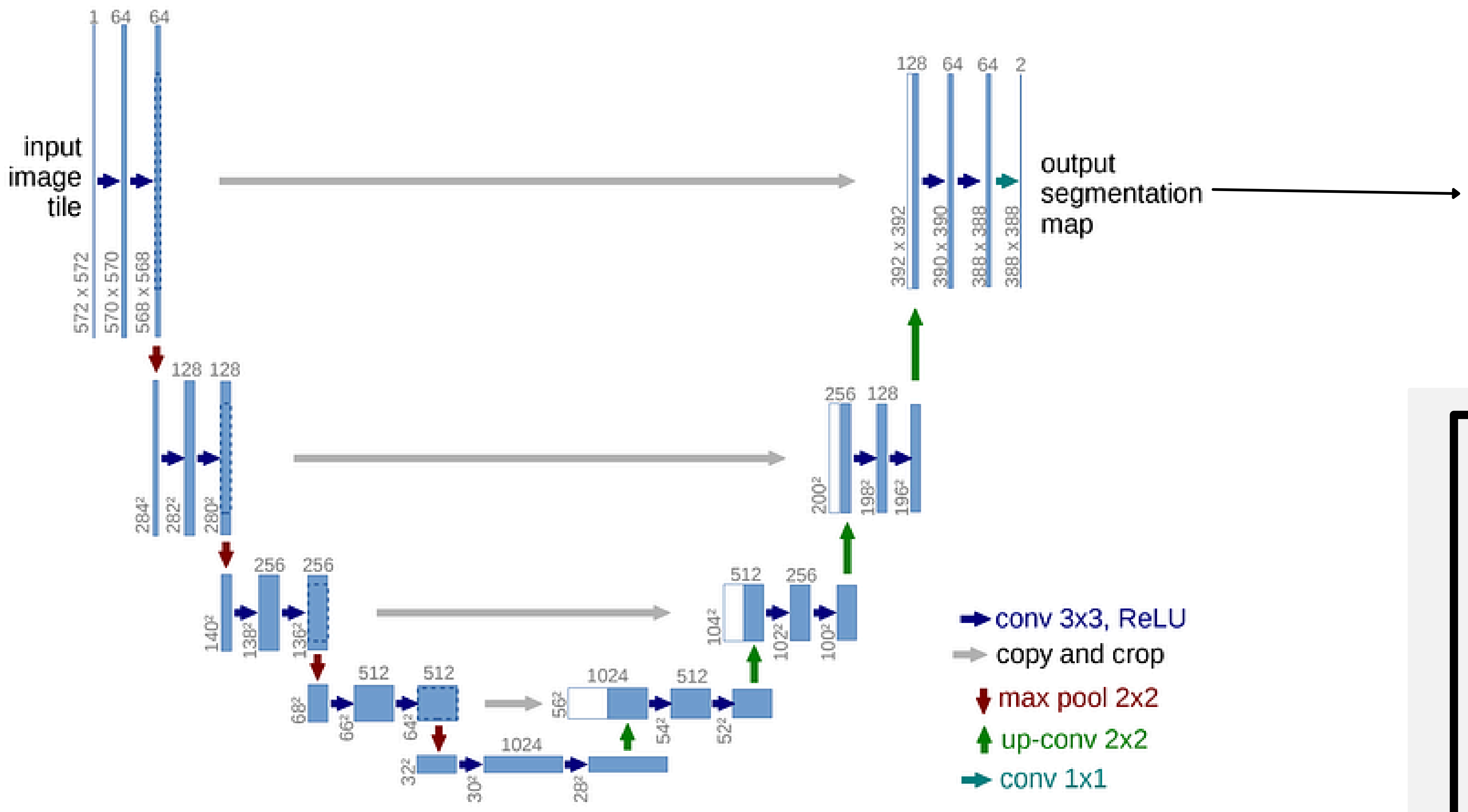


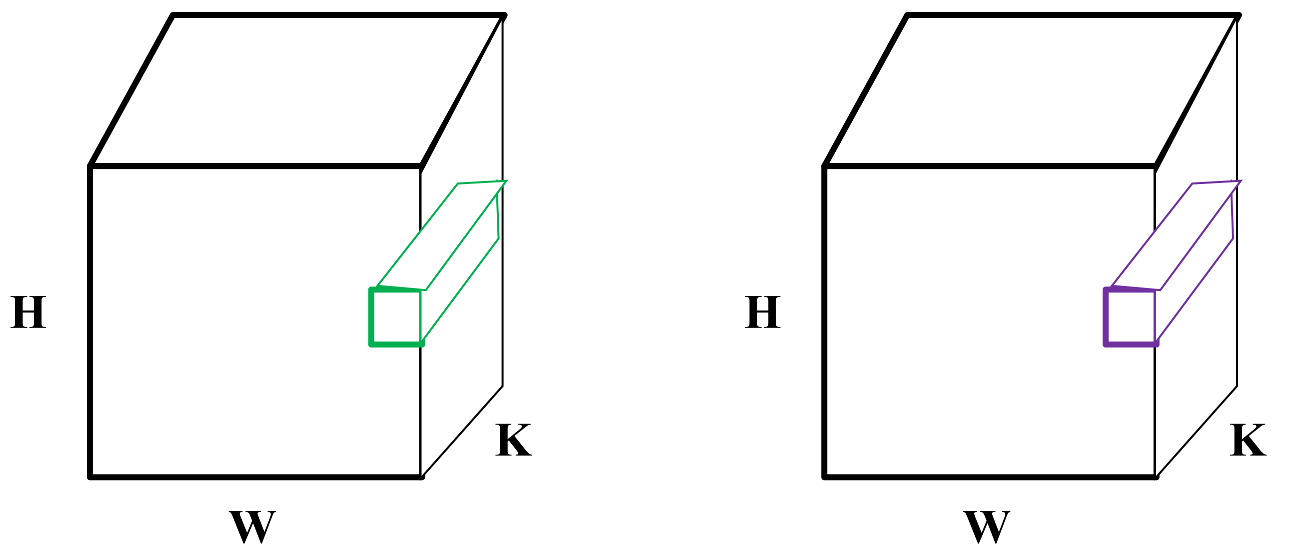
Cross Entropy



03 U-Net. Training

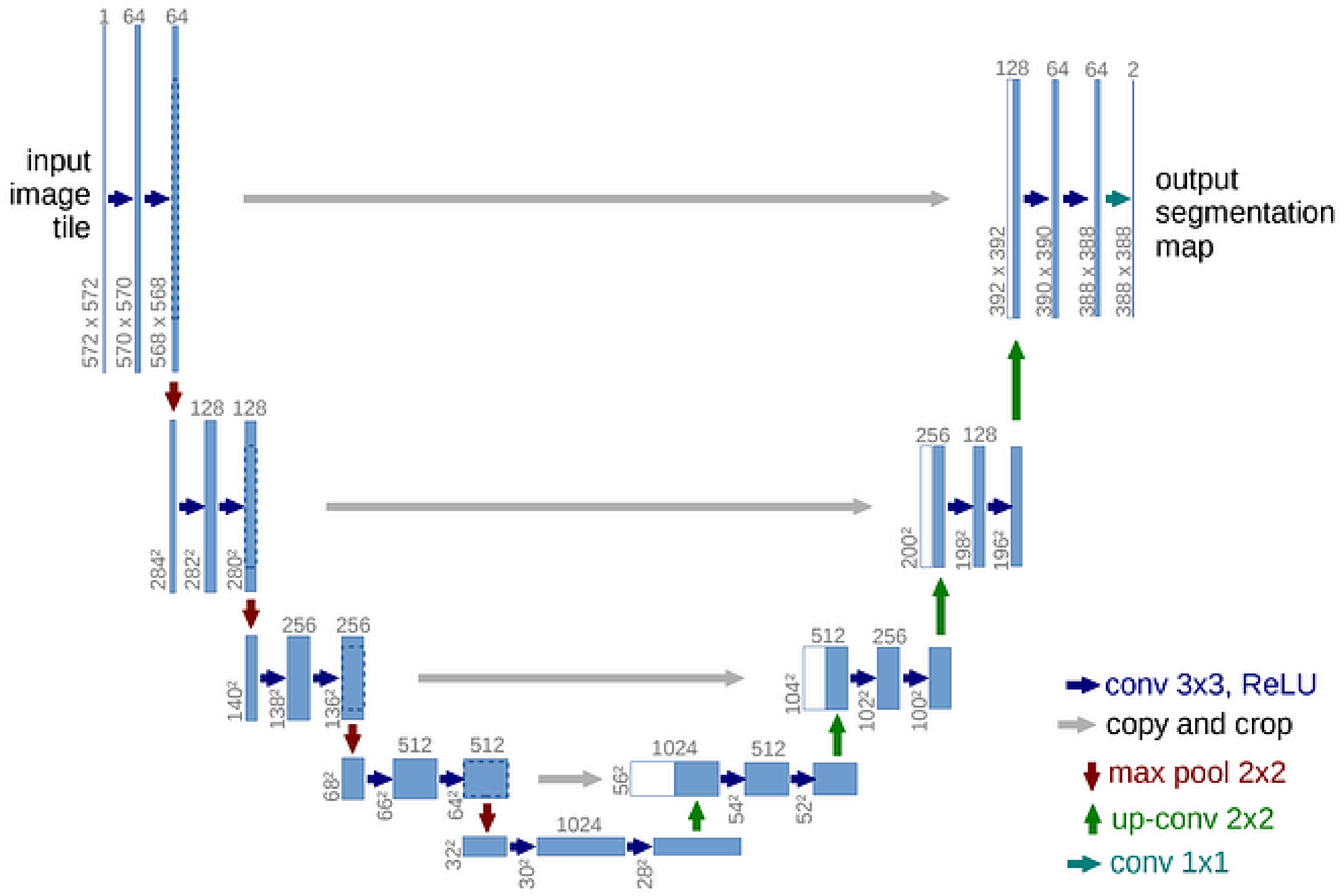
BrAIn Lab.
Brain and Artificial Intelligence Lab.

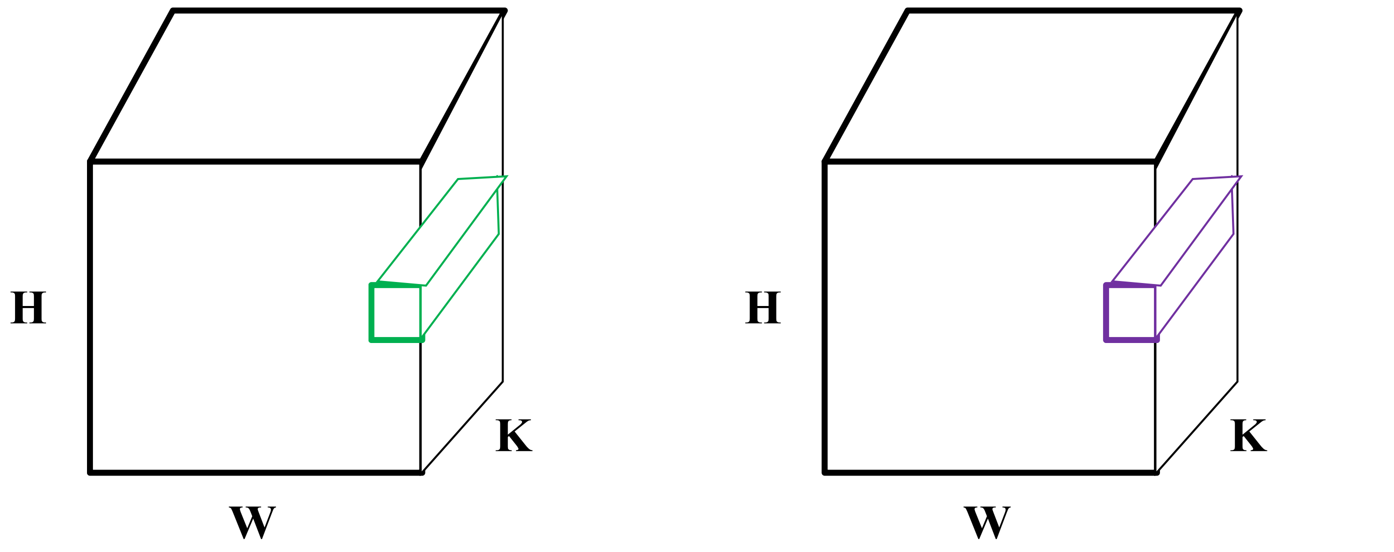




Concept:
"weight map + cross entropy"

03 U-Net. Training



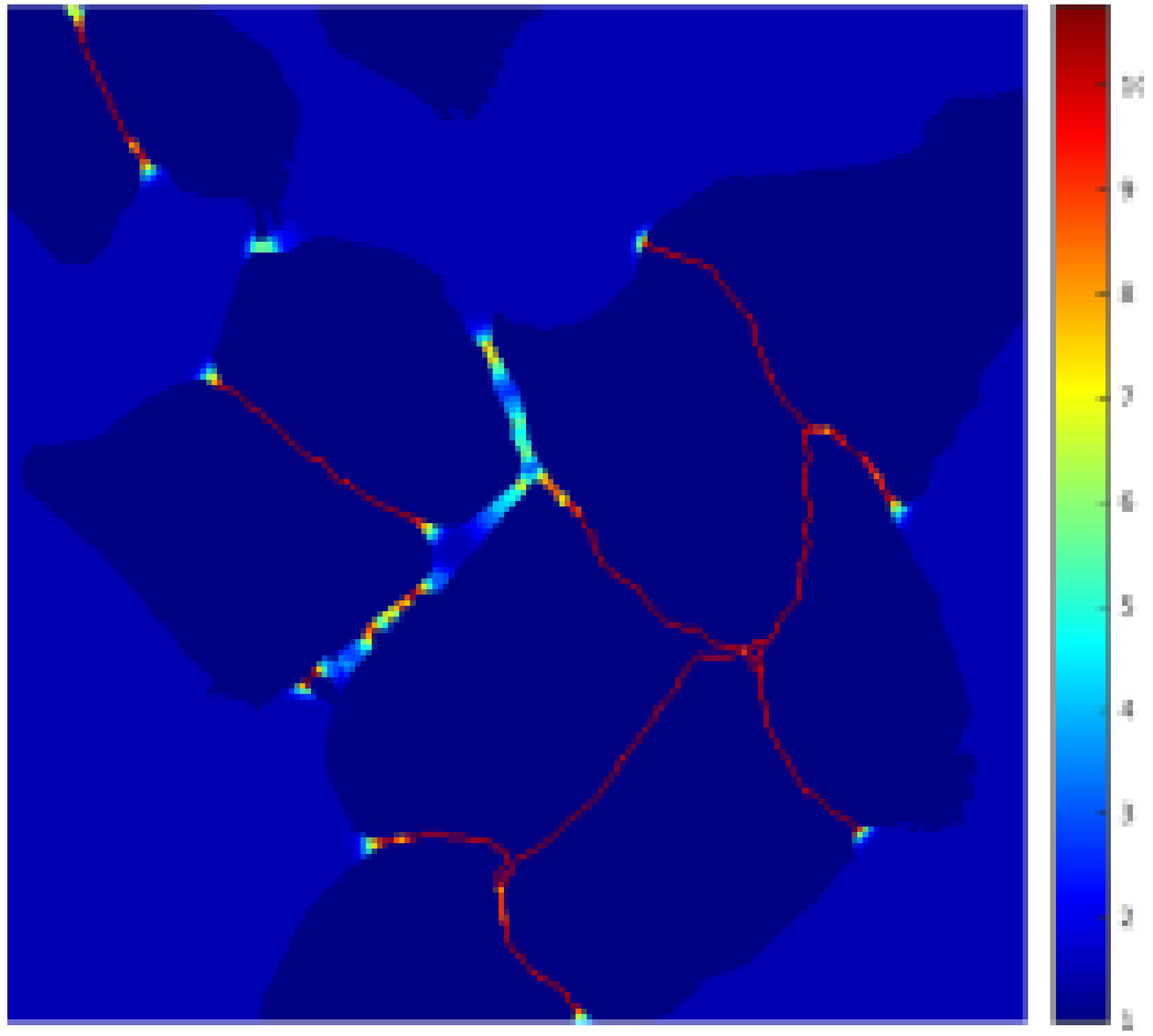


- x: specific pixel coordinate
- w(x): weighted map
- true class probability in x

03 U-Net. Training

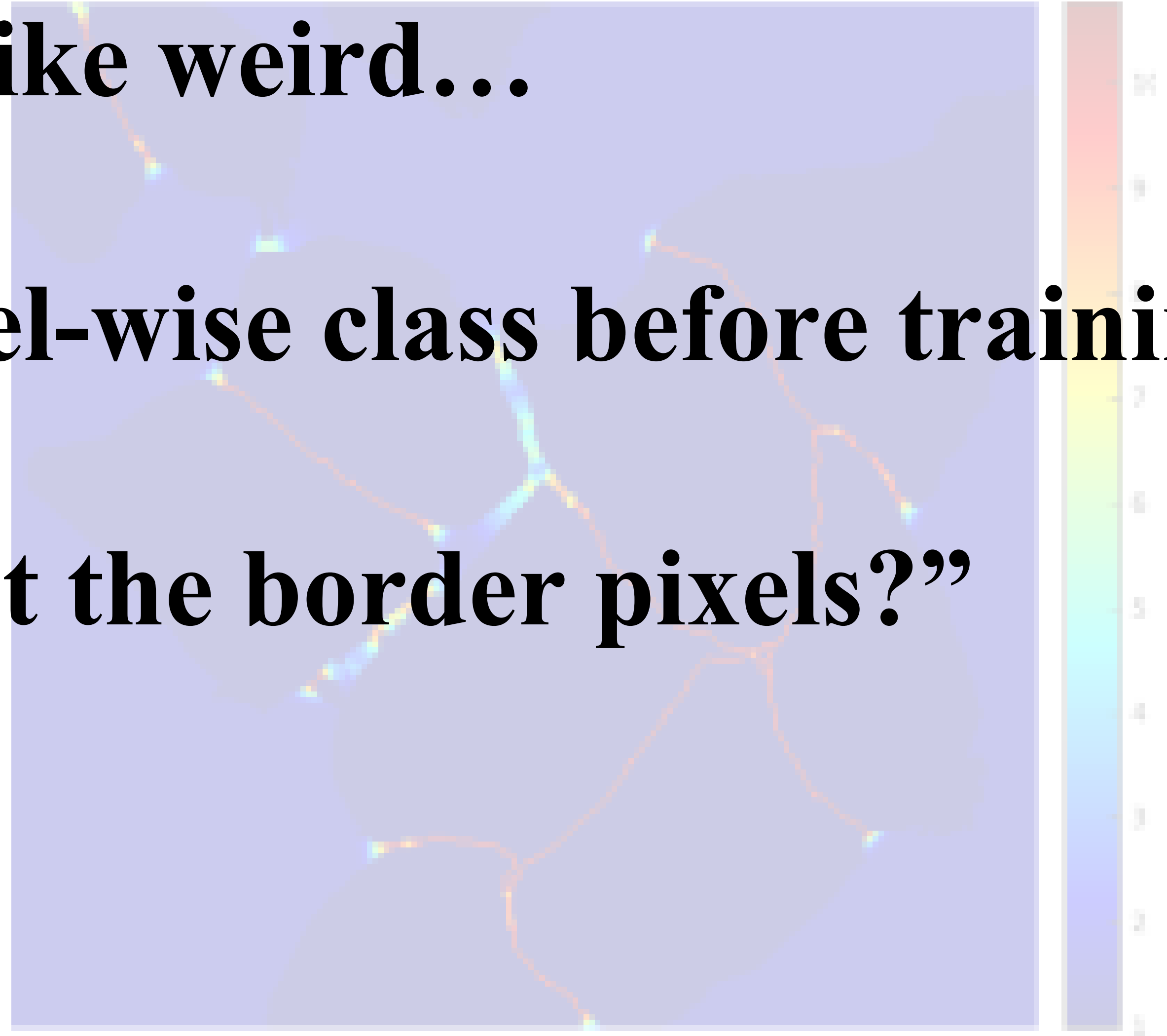
BrAIn Lab.
Brain and Artificial Intelligence Lab.

“For class imbalance and importance of segmentation border”



03 U-Net. Training

"For class imbalance and importance of segmentation border"



Weight map looks like weird…

"How we already know about the pixel-wise class before training?"

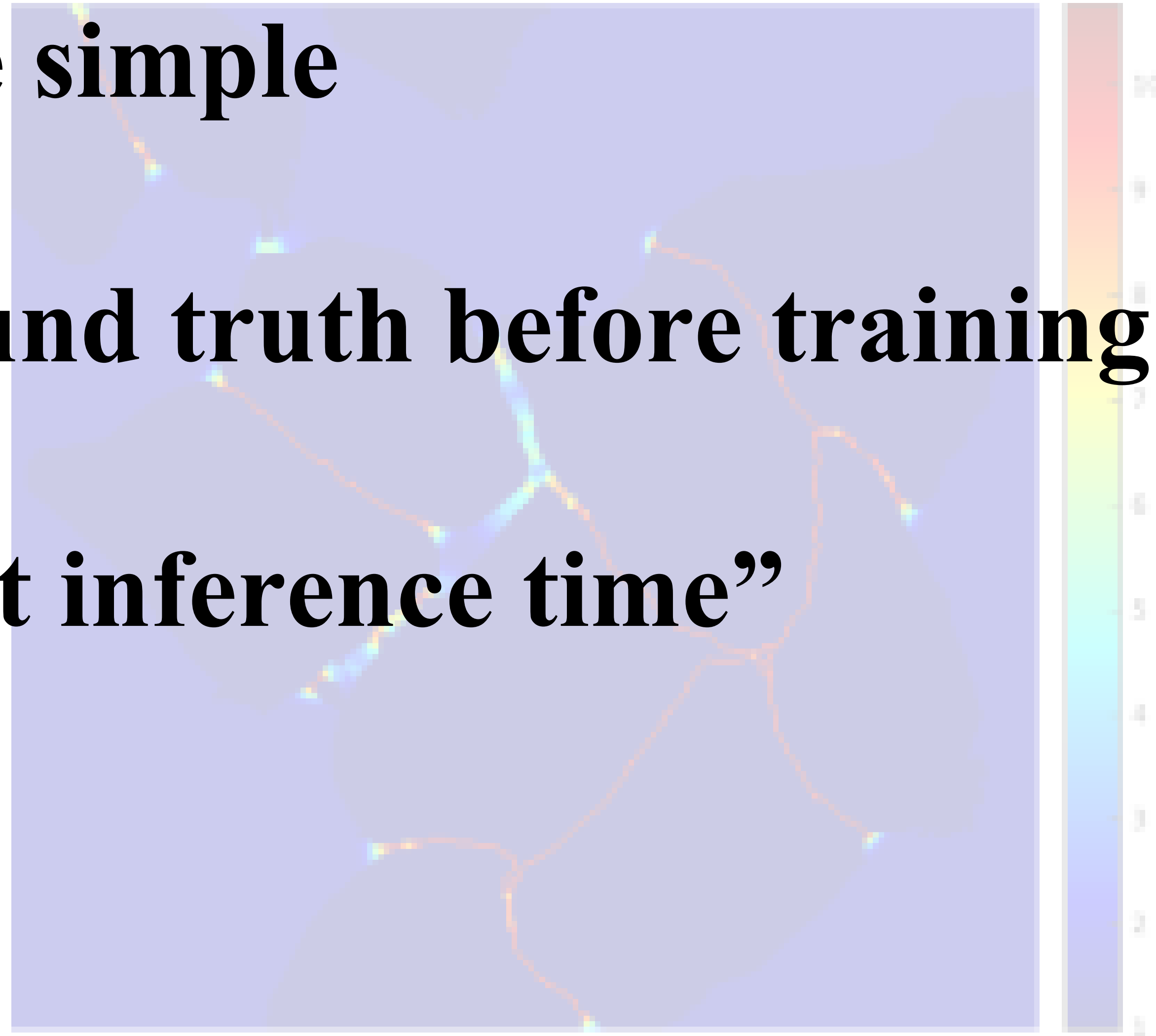
+

"How we already know about the border pixels?"

03 U-Net. Training

BrAIn Lab.
Brain and Artificial Intelligence Lab.

"For class imbalance and importance of segmentation border"



Answer is quite simple

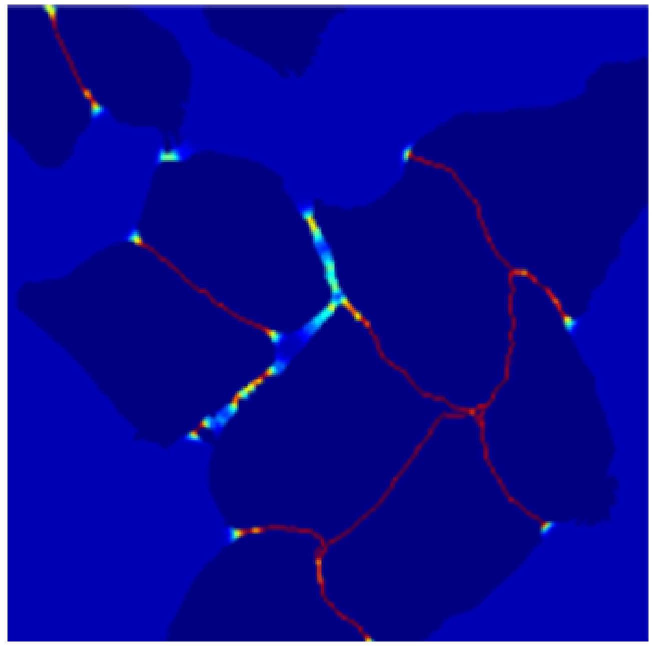
"It's already computed about ground truth before training "
+
"Don't use weight map at inference time"

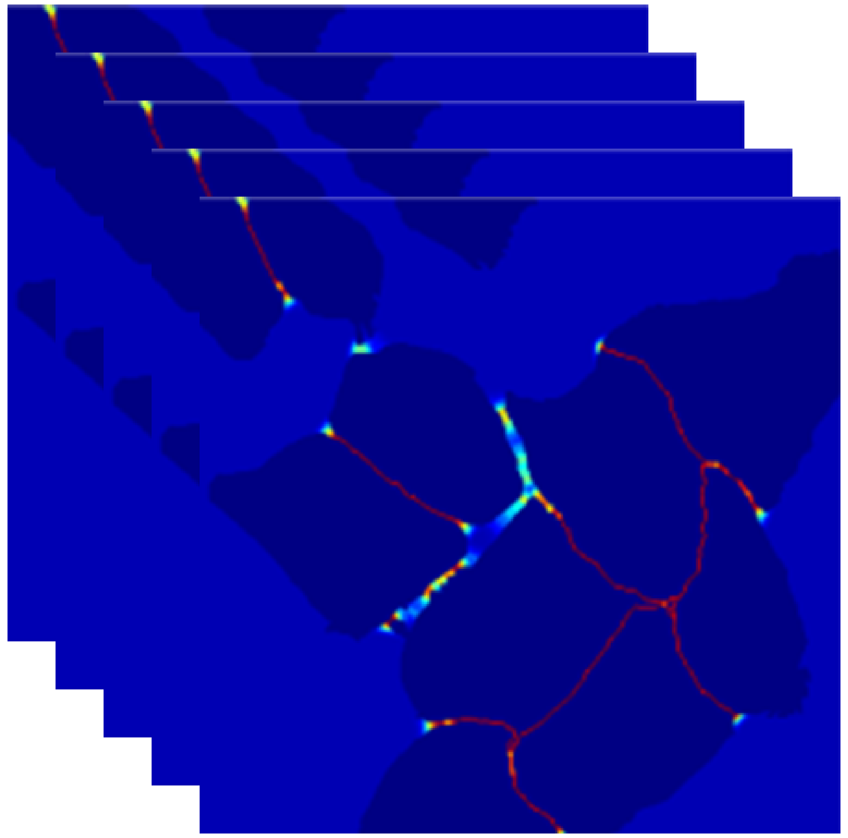
03 U-Net. Training

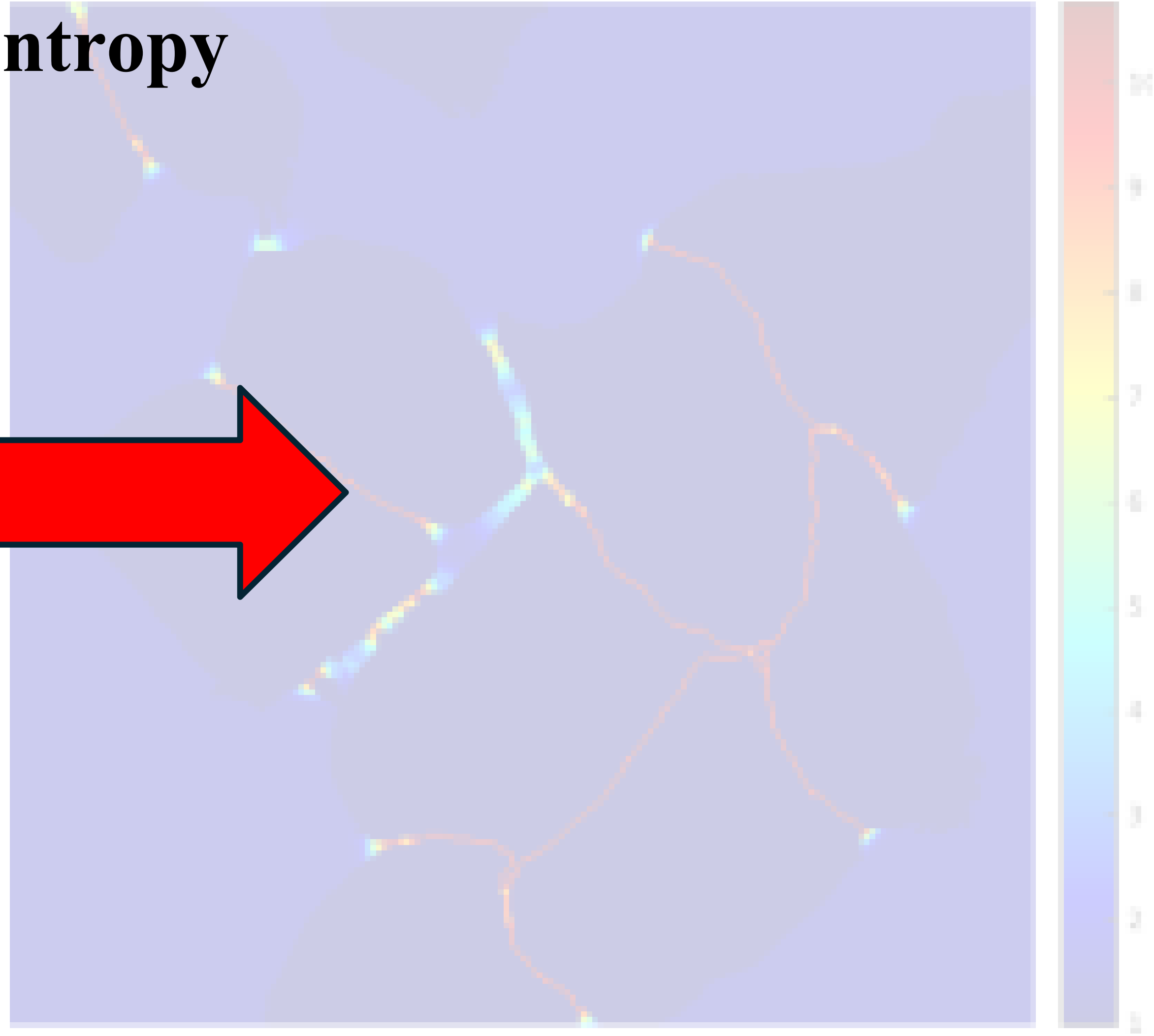
“For class imbalance and importance of segmentation border”

“How about channels?”

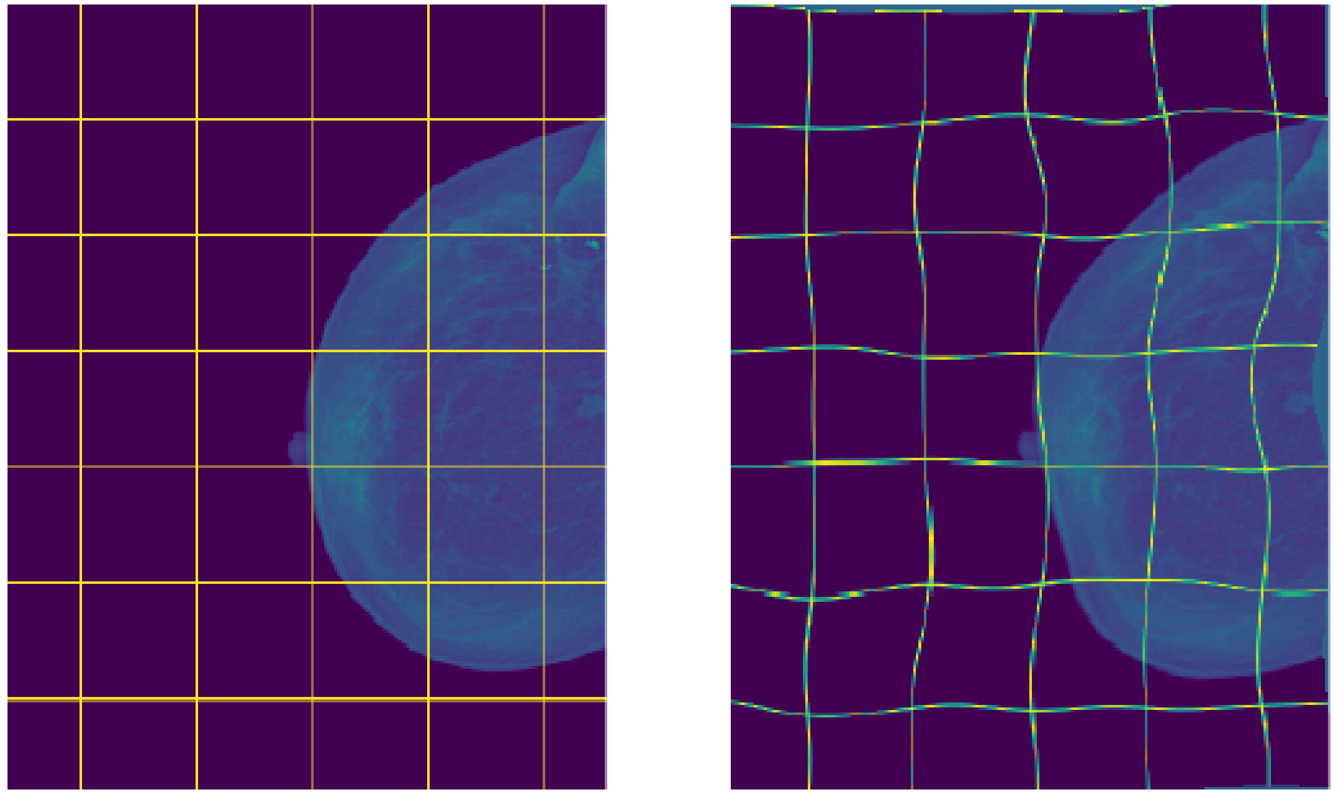
Duplicate weight map for Cross Entropy







BrAIn Lab.
Brain and Artificial Intelligence Lab.



(a) Original (b) Deformed

Fig. 3: Effects of performing elastic deformation on a mammogram.

He Initialization

Nerve cell membrane segmentation

Table 1. Ranking on the EM segmentation challenge [14] (march 6th, 2015), sorted by warping error.

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

03 U-Net. Result

Rand error analyze object separation for random two pixels

Table 1. Ranking on the EM segmentation challenge [14] (march 6th, 2015), sorted by warping error.

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

Warping error analyze cell segmentation

Analyze model output per pixel with ground truth

03 U-Net. Result

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Table 1. Ranking on the EM segmentation challenge [14] (march 6th, 2015), sorted by warping error.

?

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

Other teams Used post-processing method!

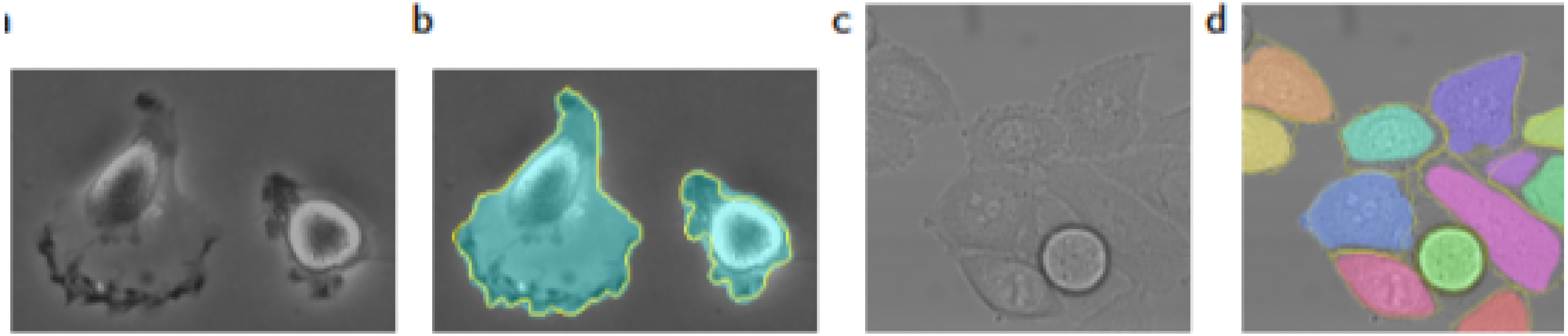
03 U-Net. Result

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Table 2. Segmentation results (IOU) on the ISBI cell tracking challenge 2015.

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

03 U-Net. Inference



Input output Input output

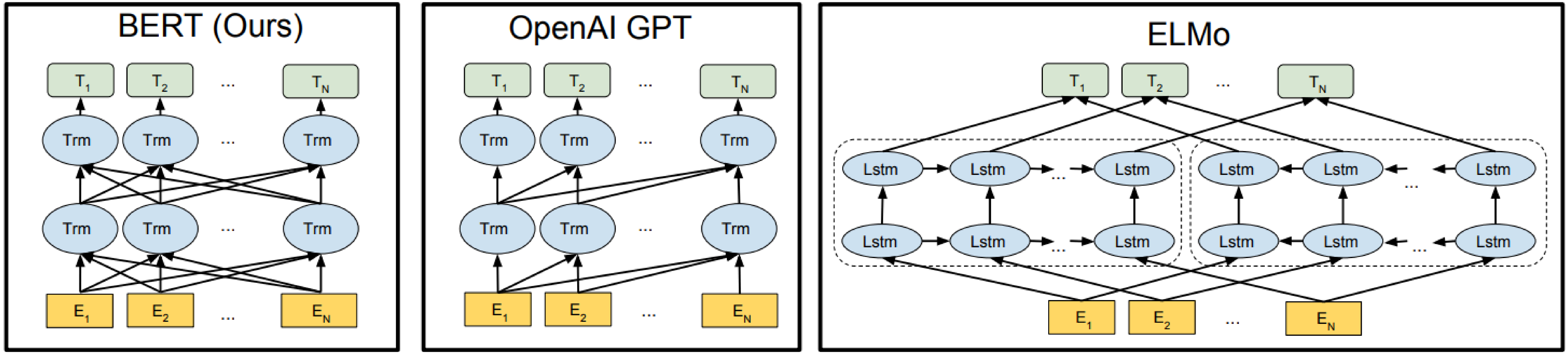
Bidirectional Encoder Representations from Transformers

Bidirectional Encoder Representations from Transformers

"Bidirectional Learning for Context Understanding"

04 BERT vs.

BrAIn Lab.
Brain and Artificial Intelligence Lab.



1. **BERT: Bidirectional learning with Using Self-attention**
2. **GPT-1: One-Directional Learning with Masking self-attention (Left-to-Right model, LTR)**
3. **ELMo: Concatenation result with “LTR LSTM model + RTL LSTM model”**

Feature-based & Fine-tuning

“Using pre-trained model(freeze) for feature extractor”

BrAIn Lab.
Brain and Artificial Intelligence Lab.

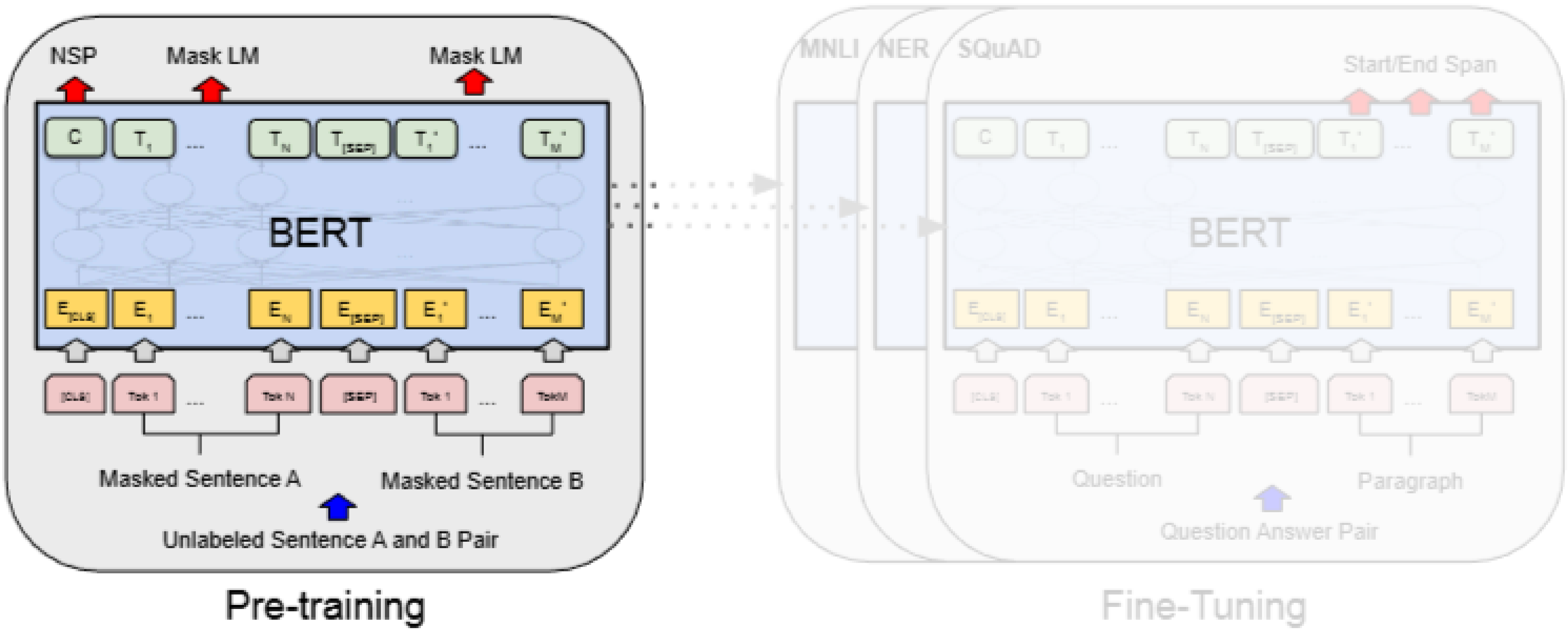
Feature-based & Fine-tuning

“transfer learning(no freeze) with pre-trained model”

Unsupervised Fine-tuning Approaches

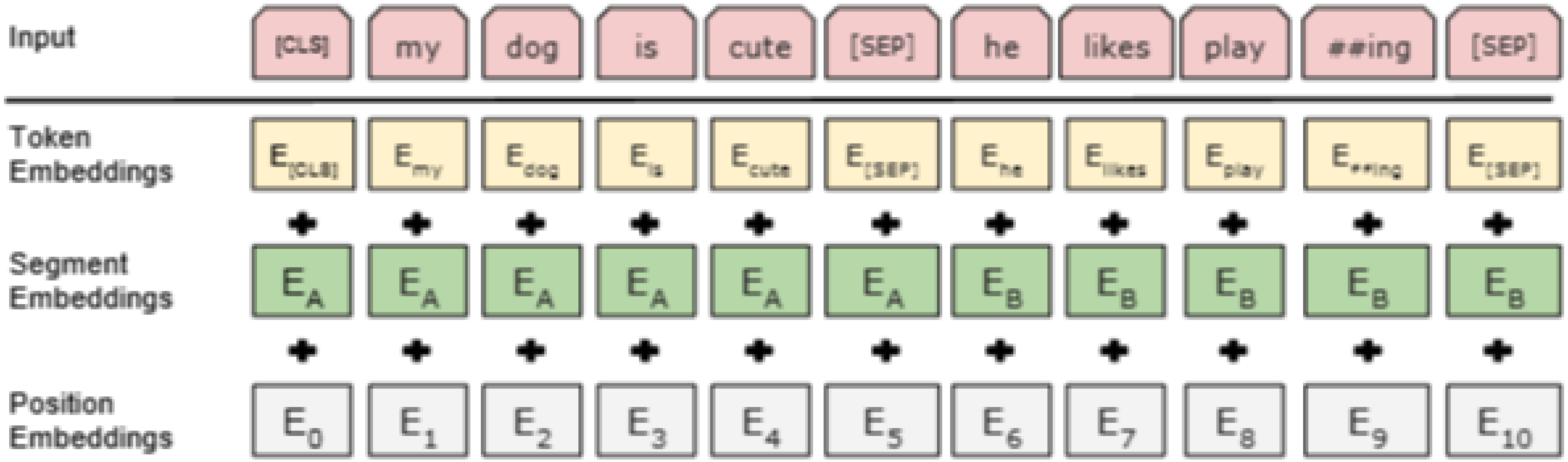
04 BERT. Overview

BrAIn Lab.
Brain and Artificial Intelligence Lab.



04 BERT. Input

BrAIn Lab.
Brain and Artificial Intelligence Lab.



1. "Segment Embeddings + Separate Token" works about separating sentences

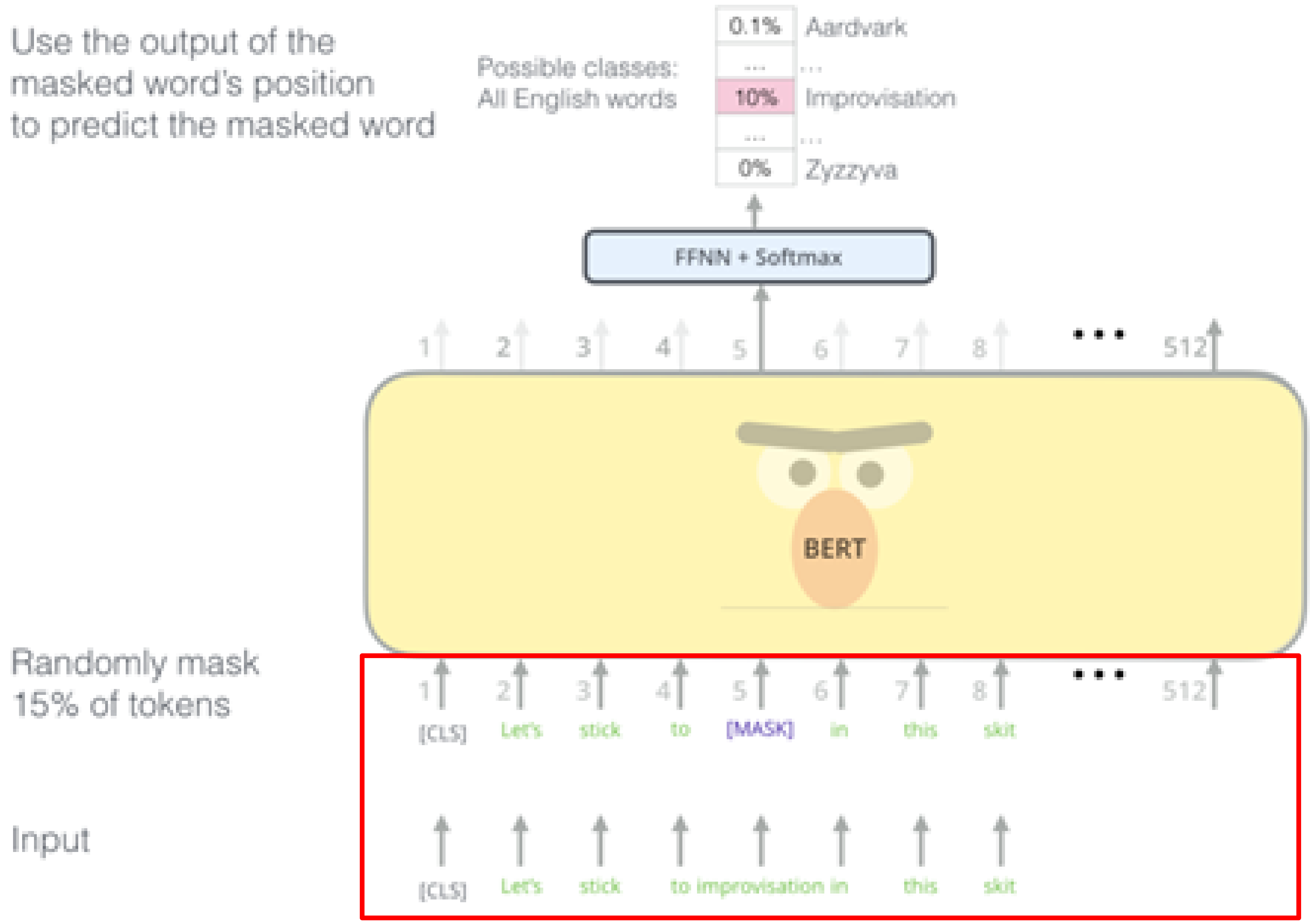
2. Class token have whole context info for whole sentences (because of self-attention and no info itself)

3. Sum of embeddings info use for input

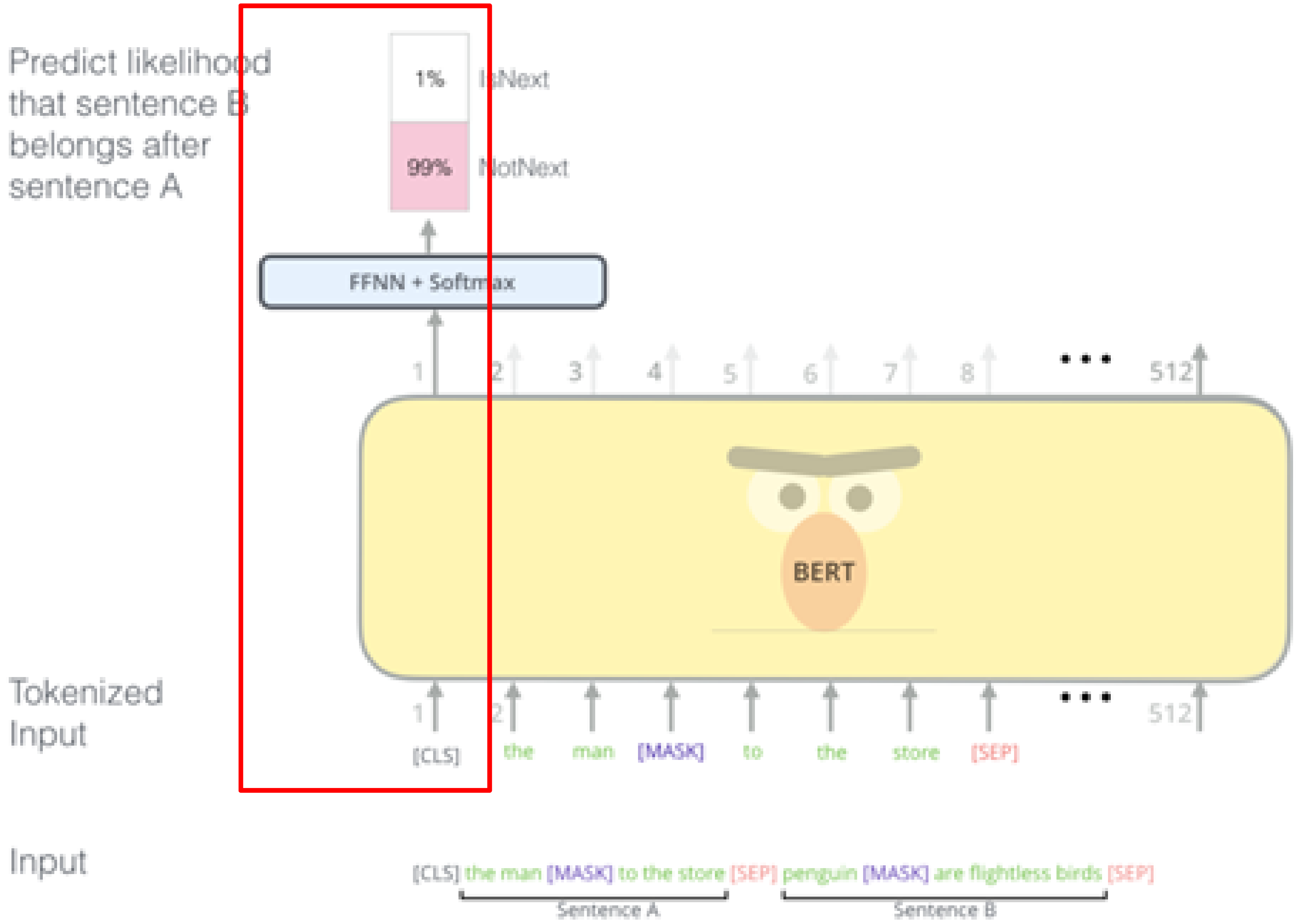
04 BERT. pre-training

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Class token use for Next Sentence Prediction & classification



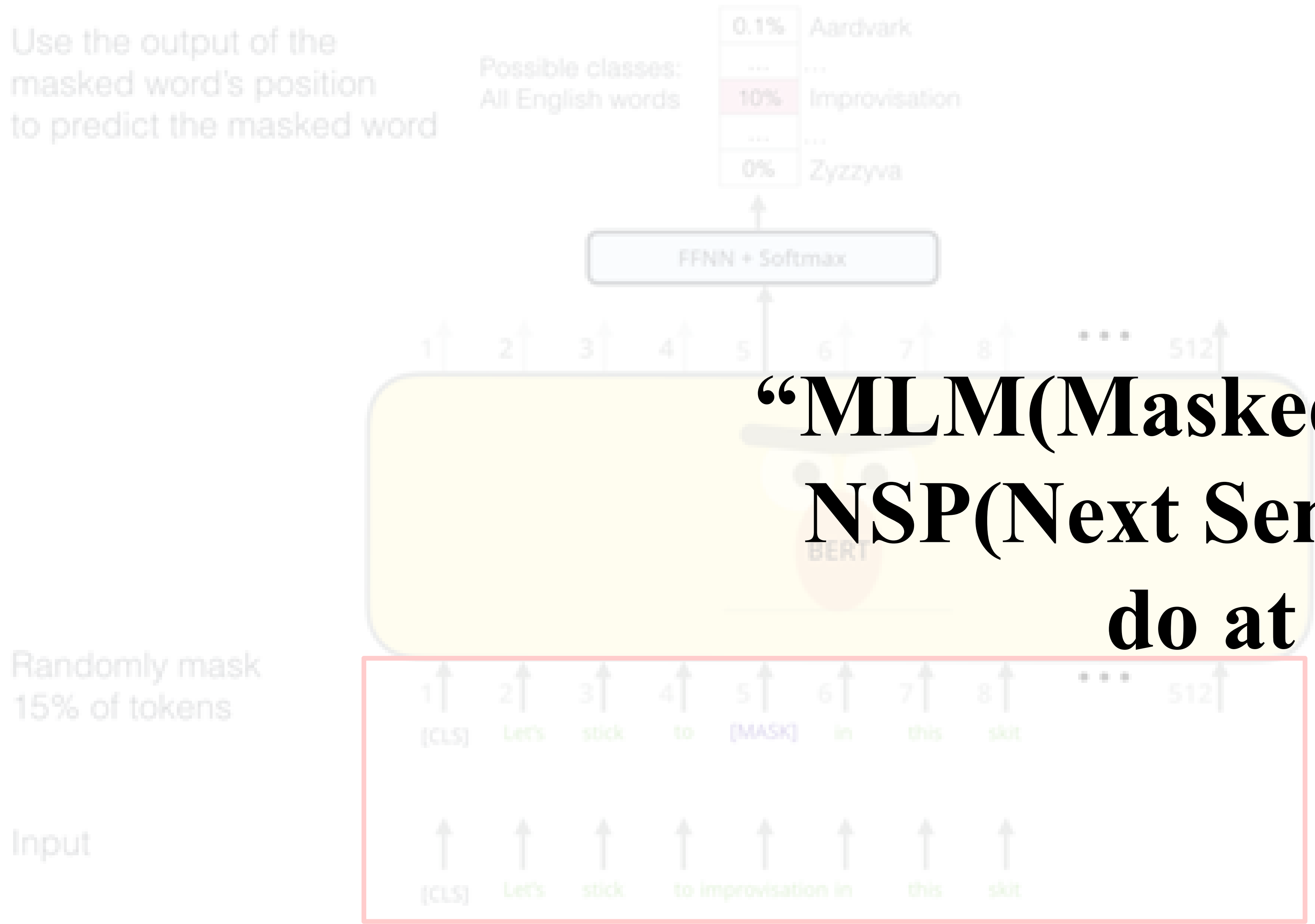
Mask Random Token

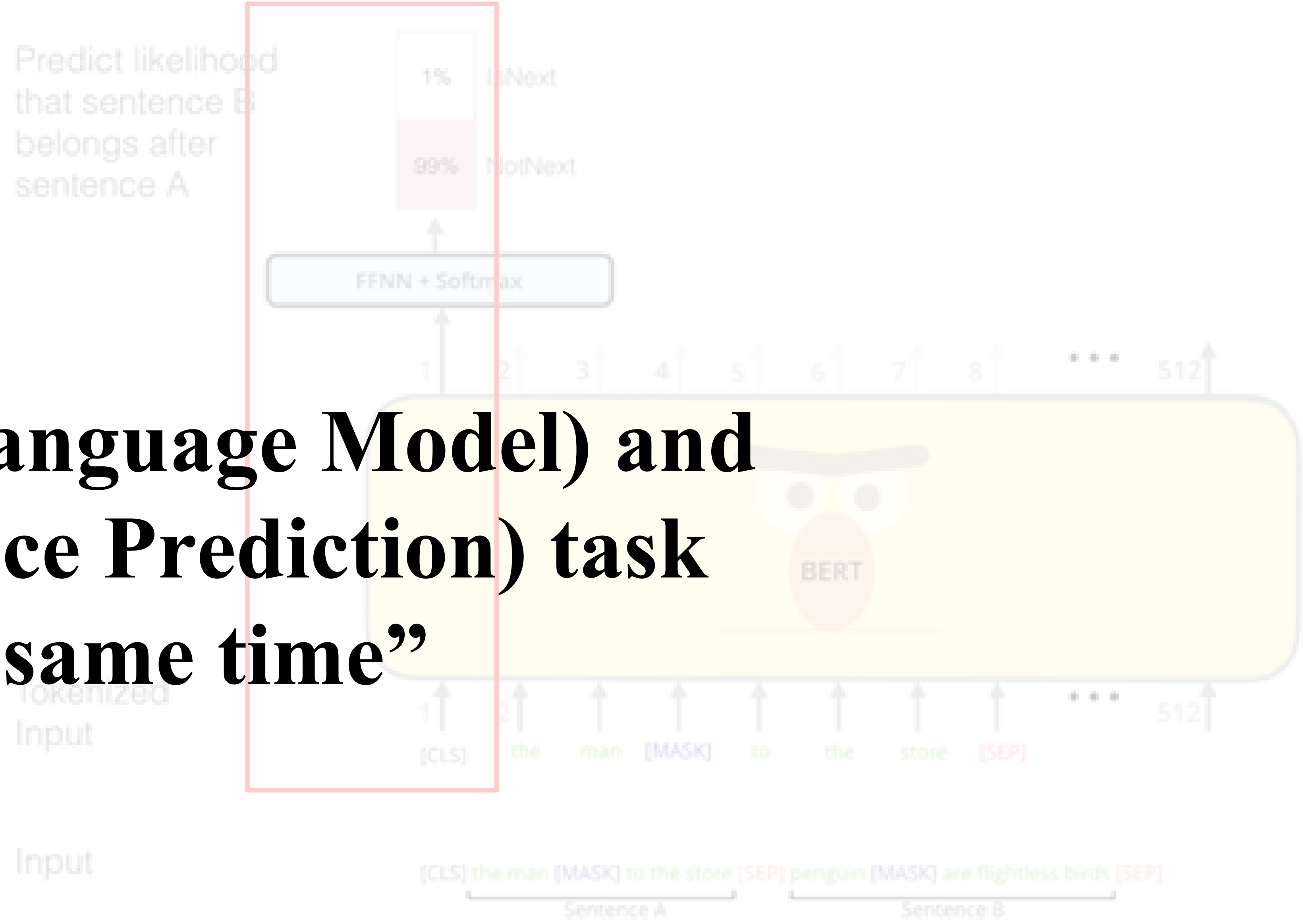


04 BERT. pre-training

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Class token use for Next Sentence Prediction & classification





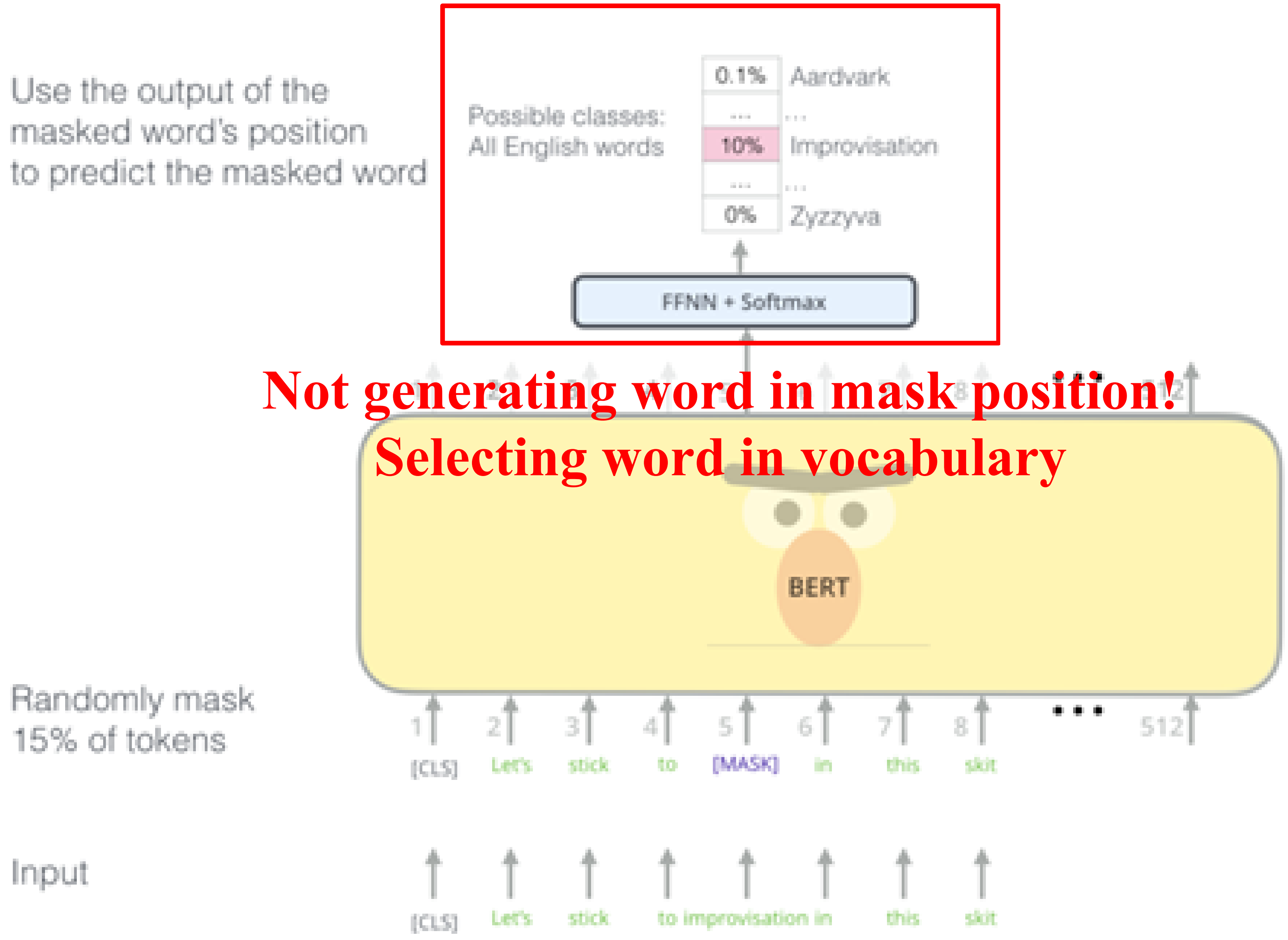
"MLM(Masked Language Model) and NSP(Next Sentence Prediction) task do at the same time"

Mask Random Token

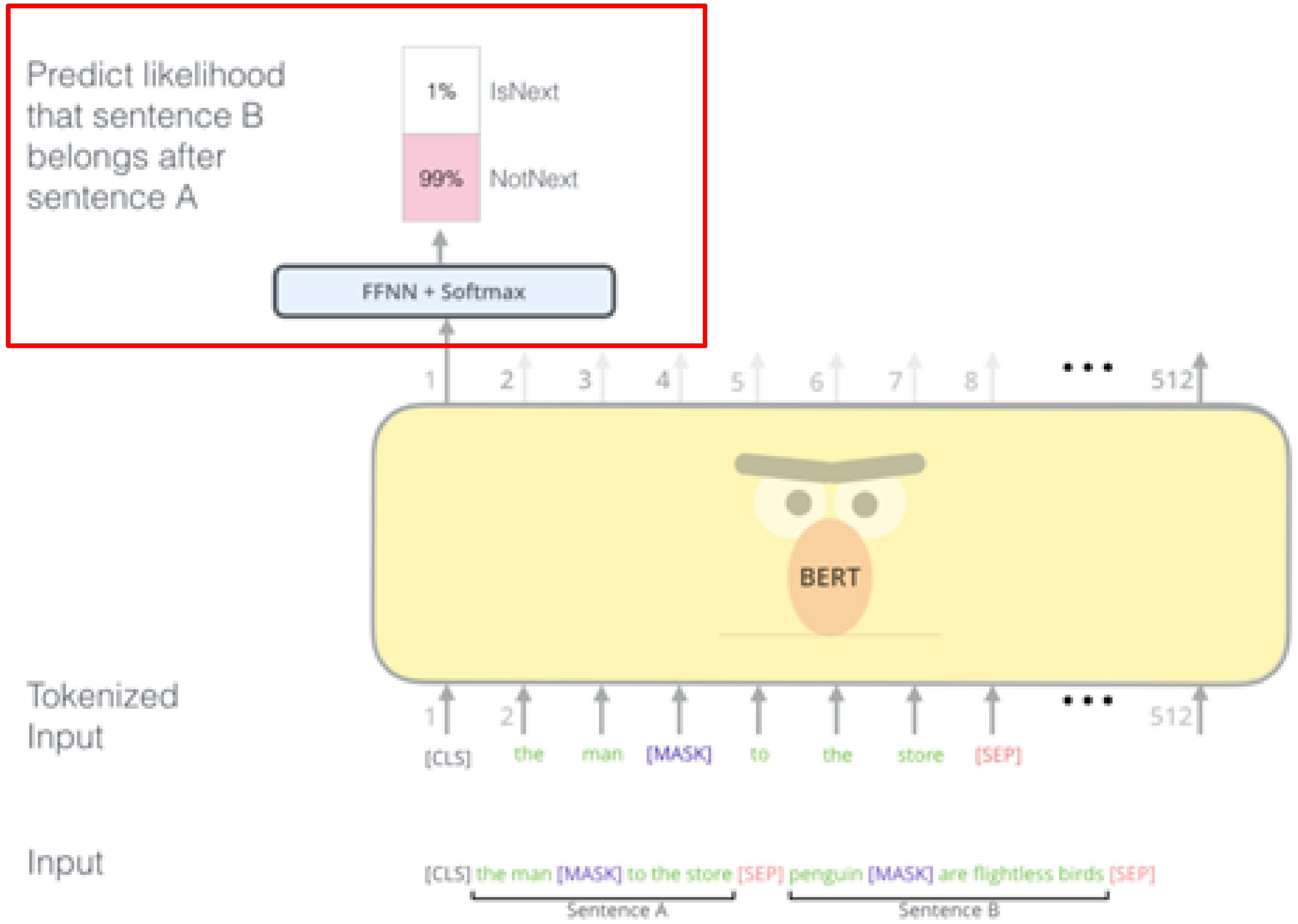
04 BERT. pre-training

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Vocabulary and NSP Answer already exist before pre-training

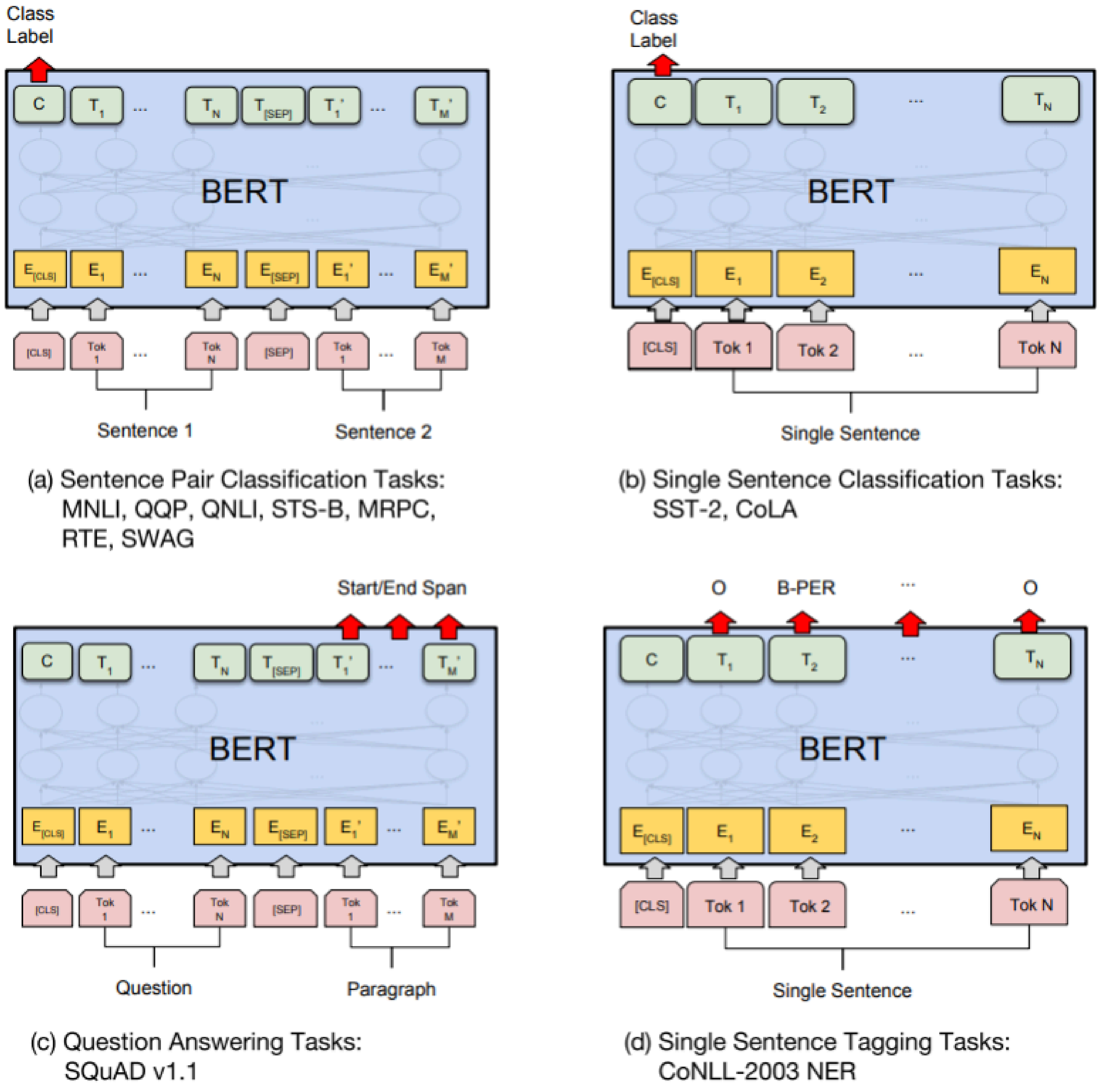


Not generating word in mask position!
Selecting word in vocabulary



04 BERT. Fine-tuning

BrAIn Lab.
Brain and Artificial Intelligence Lab.



1. **A separate fine-tuned model is required for each task**
2. **In task (a), the model determines whether two sentences are semantically identical**
3. **In task (b), the model matches or classifies the overall properties or characteristics of an entire sentence**
4. **In task (c), the model is given a passage along with a question and must predict the span in the passage where the correct answer appears**
5. **In task (d), the model classifies the meaning or role of each token, such as whether a word is a verb, a noun, a person's name and so on.**

04 BERT. Result

BrAIn Lab.
Brain and Artificial Intelligence Lab.

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	Average
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
$\text{BERT}_{\text{BASE}}$	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
$\text{BERT}_{\text{LARGE}}$	**86.7/85.9**	**72.1**	**92.7**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**82.1**

Thank You

2026.00.00

BrainLAB Journal Club
Department of Applied Artificial Intelligence
Jeong SangYeop