Multi-Scale Explainable Feature Learning for Pathological Image Analysis using Convolutional Neural Networks

National Institute of Advanced Industrial Science and Technology (AIST), Japan

○Kazuki Uehara, Masahiro Murakawa, Hirokazu Nosato, Hidenori Sakanashi

Contributions

We propose an **explainable** feature learning method for CADs of pathological images using a CNN:

- Extracting **multi-scale features**
- Constructing **feature dictionaries** with vector quantization
- **Visualizing items** in the dictionaries as images

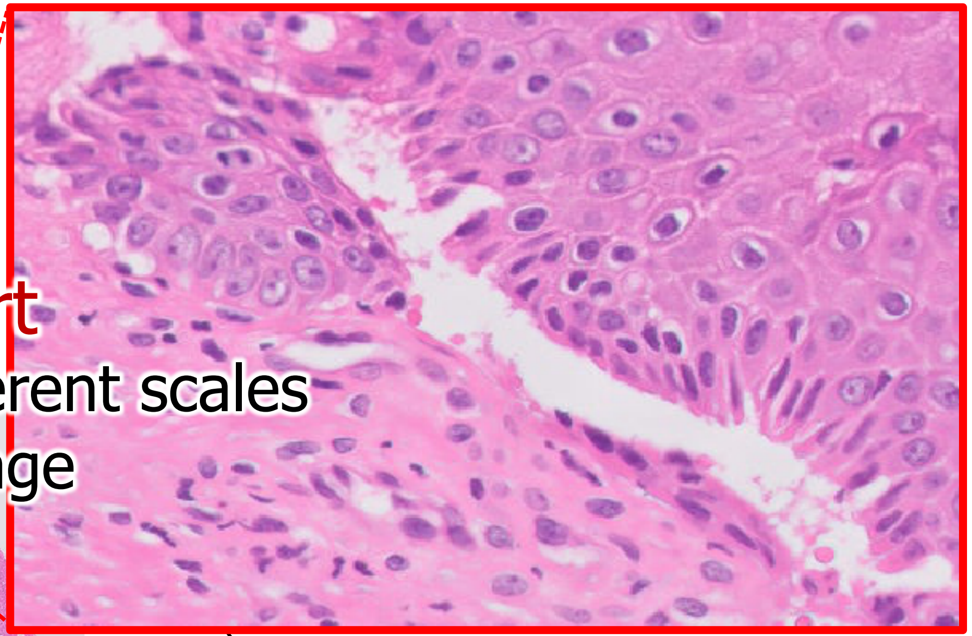
Experimental results showed 0.89 of AUROC on detecting atypical tissues in pathological images

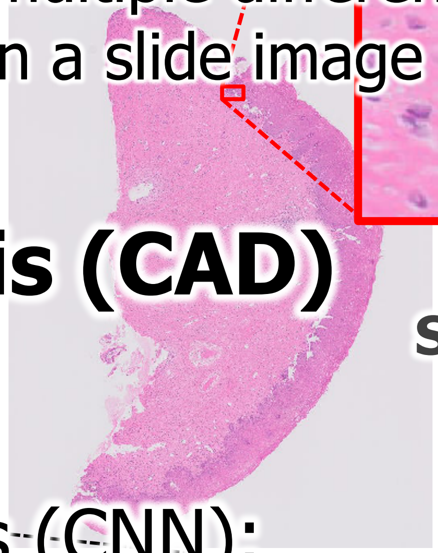
Introduction

Pathology

- To determine treatment of cancer
- Requiring considerable time and effort

  Exploring tiny atypical cells at multiple different scales
  by evaluating dozens of cells on a slide image





Computer Aided Diagnosis (CAD)

Relieve pathologist’s burden

Convolutional Neural Networks (CNN):

High accuracy for pathological image analysis

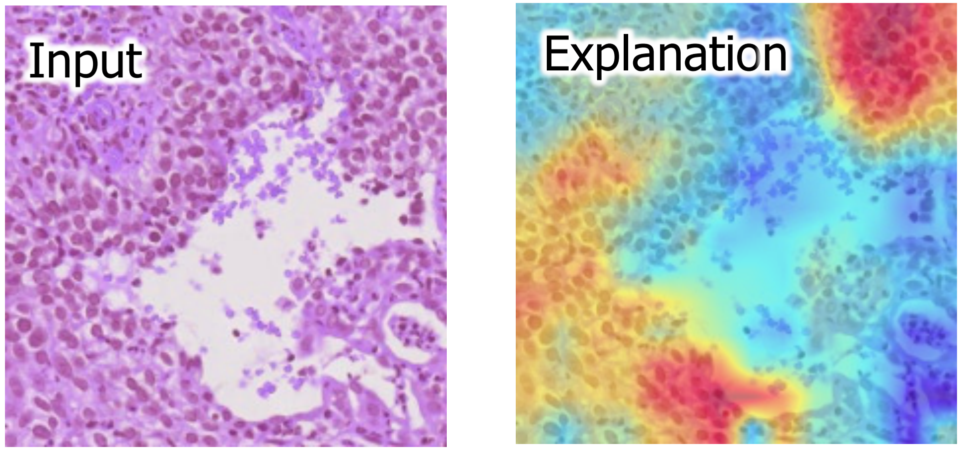
Explainability of CNNs

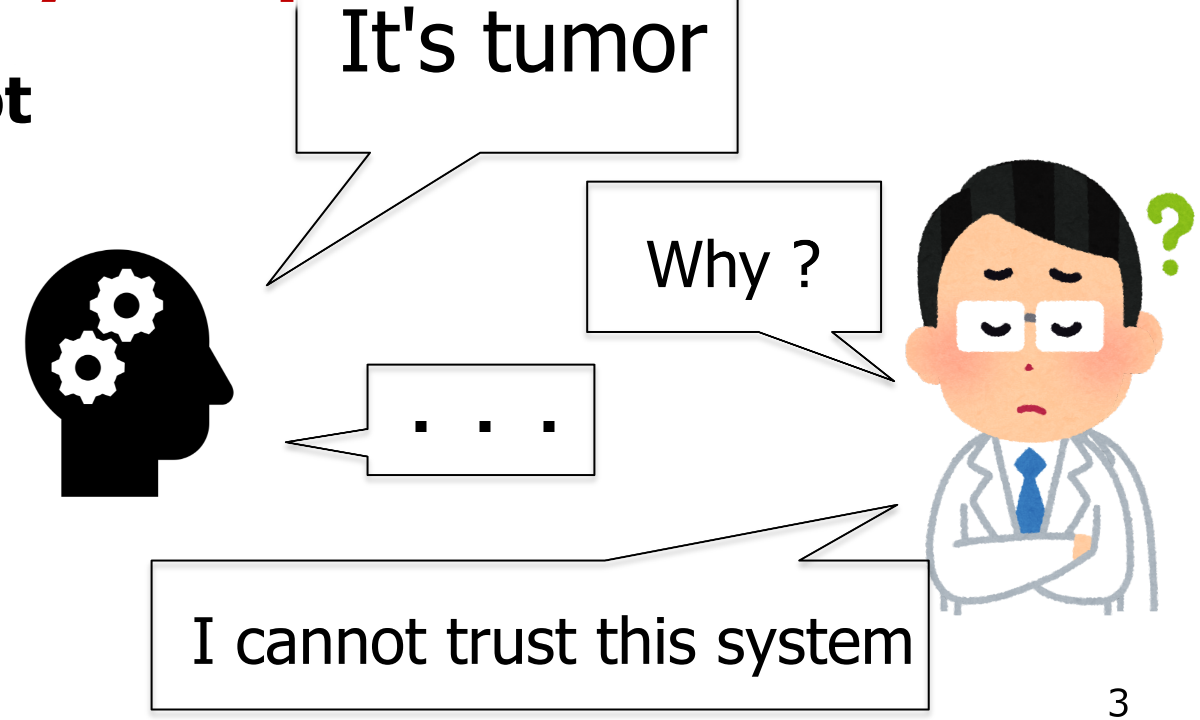
CAD systems should be accurate and explainable to ensure their reliability

Explainability: Basis of diagnoses can be interpreted by humans

Decisions made by CNNs are hardly interpretable

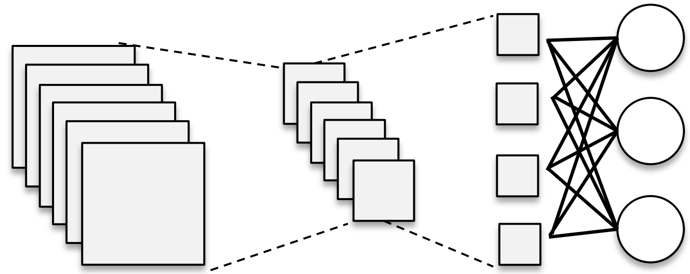
Activation based explanations cannot tell the reasons for their decisions



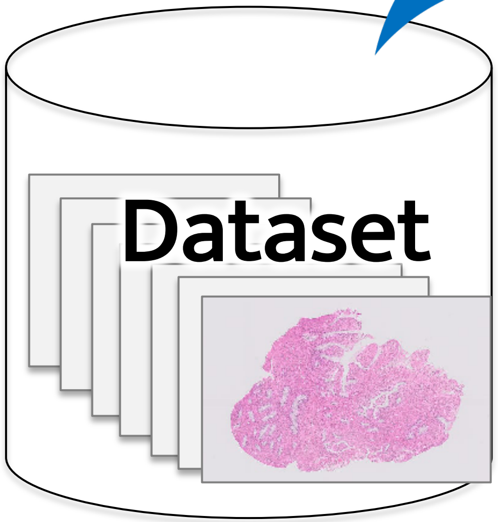


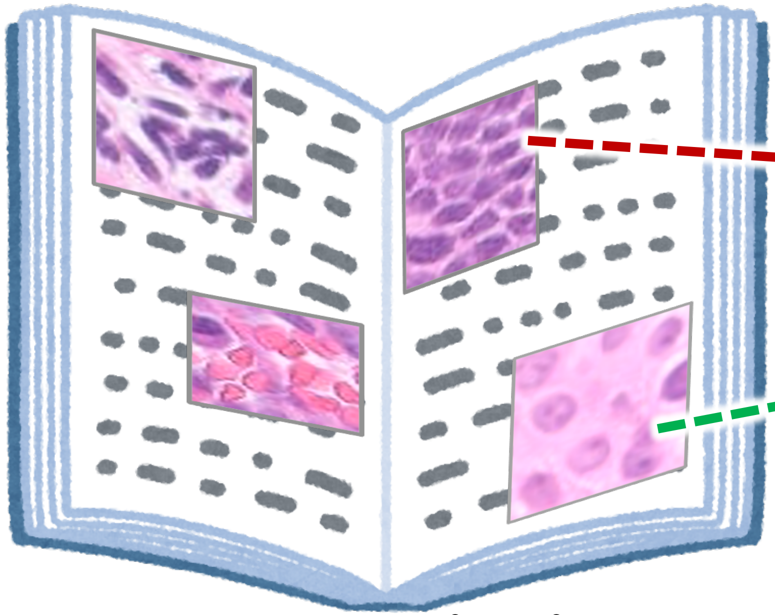
Objective

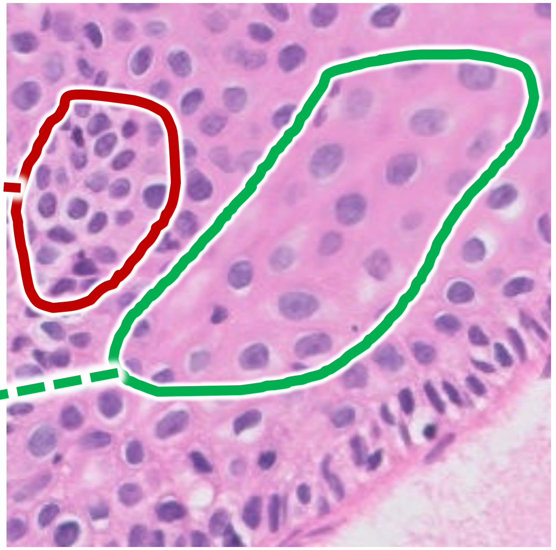
Accurate and Explainable diagnosis based on a dictionary using a CNN



Important features
for diagnosis

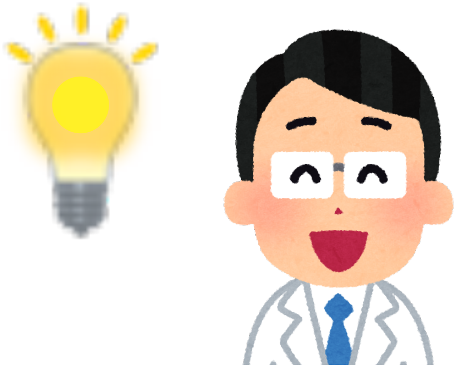




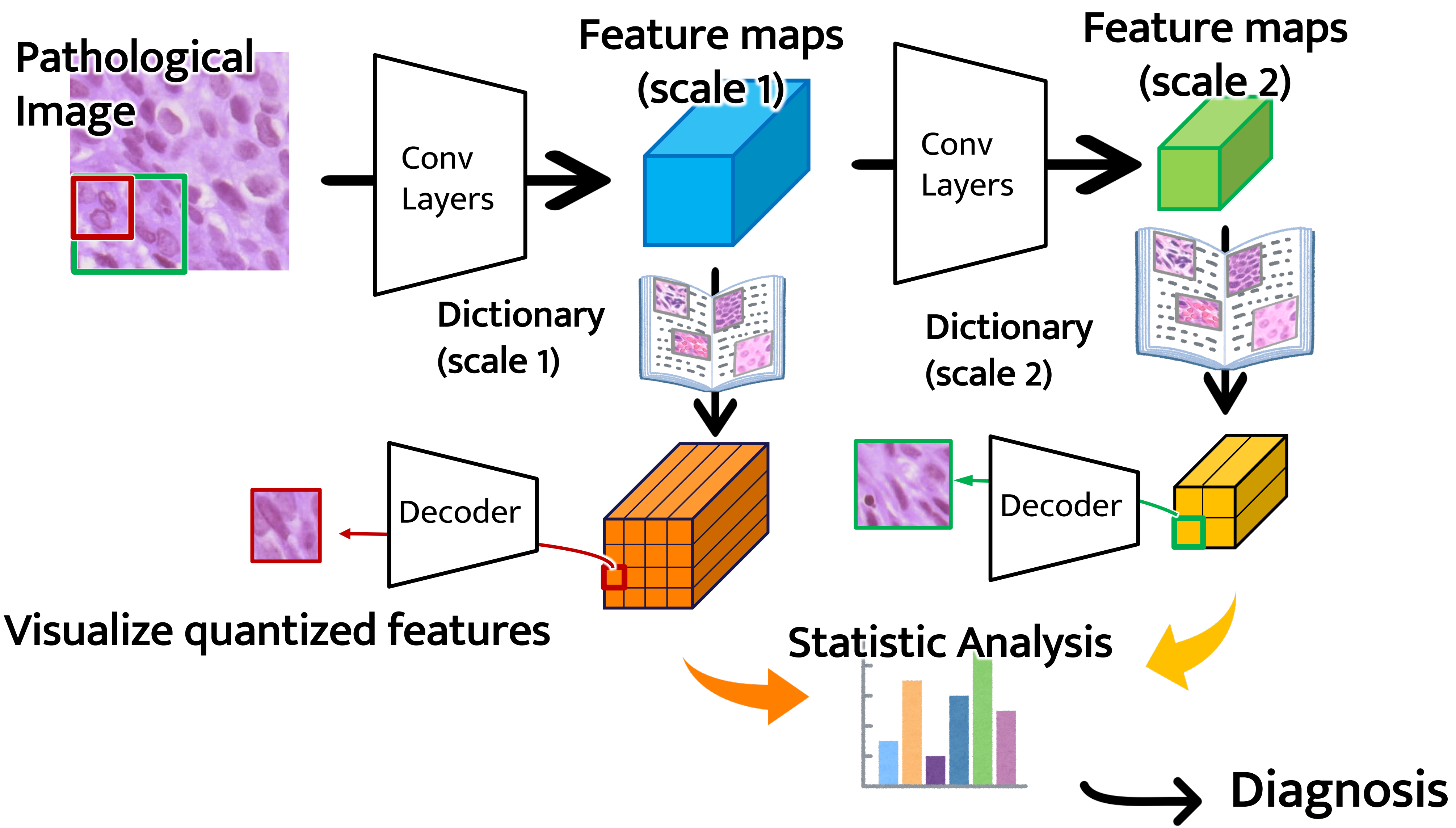


Feature Dictionary

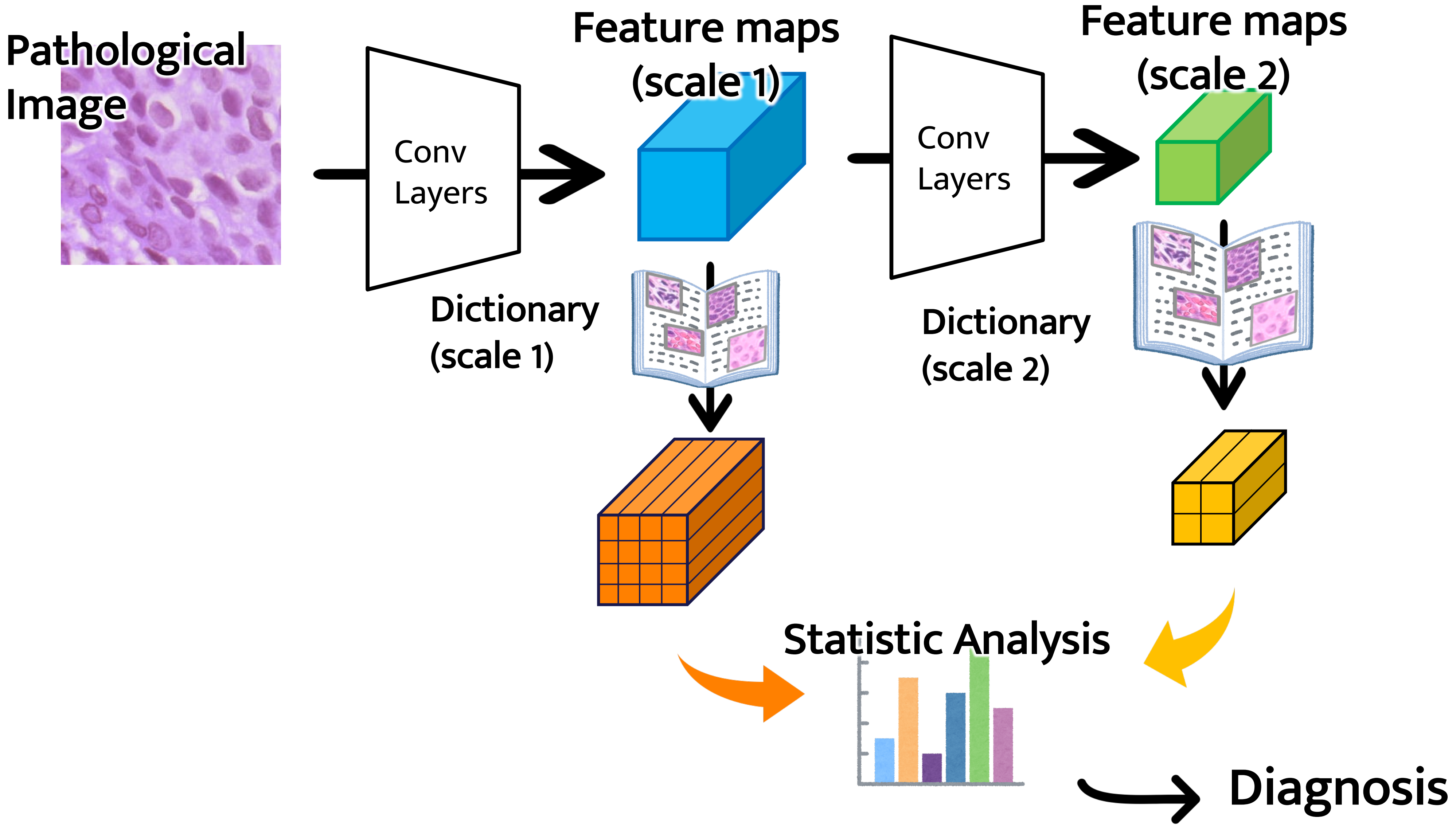
- Clinical Findings
- Appearance



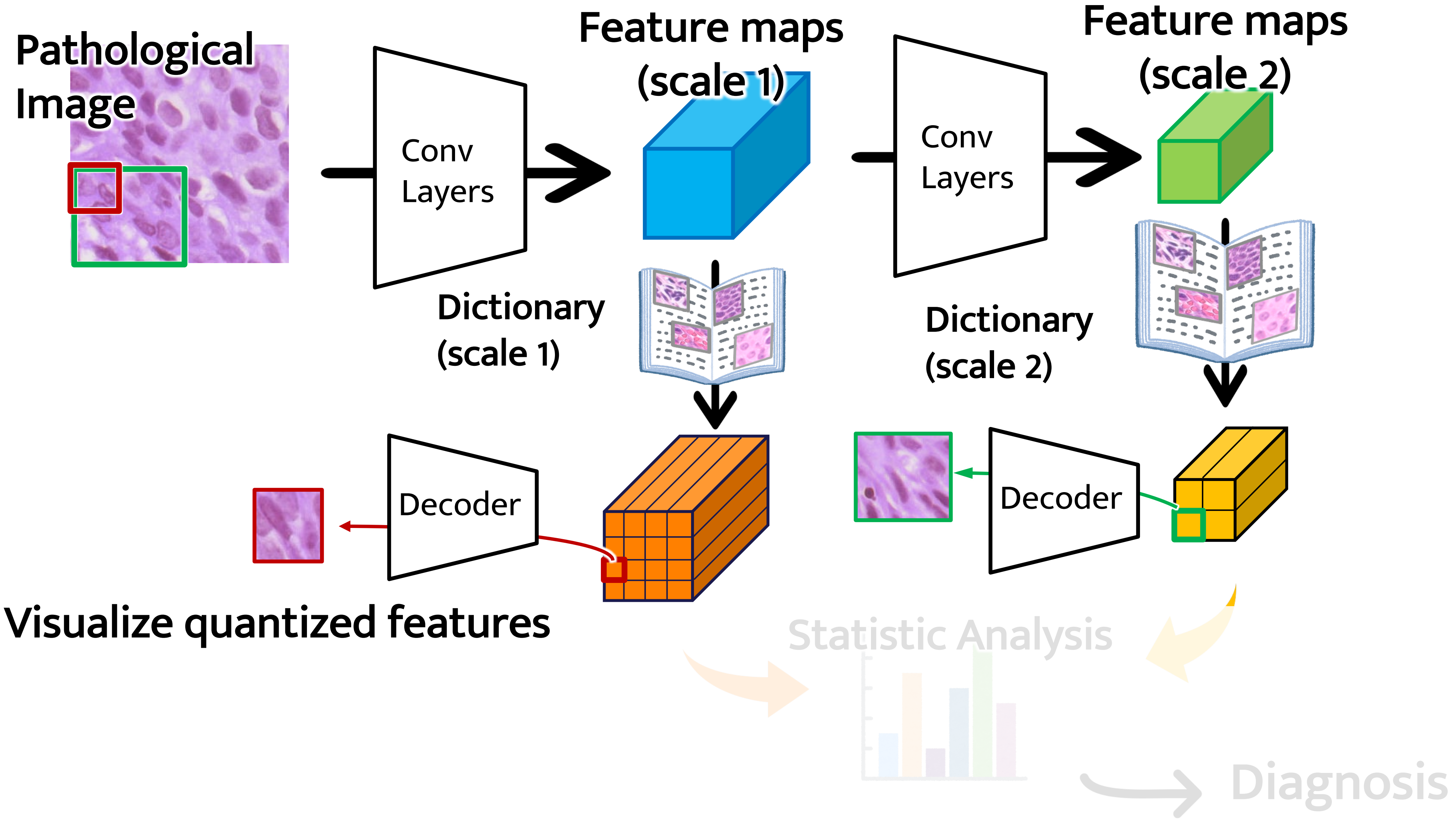
Multi-scale Explainable Feature Learning



Multi-scale Explainable Feature Learning

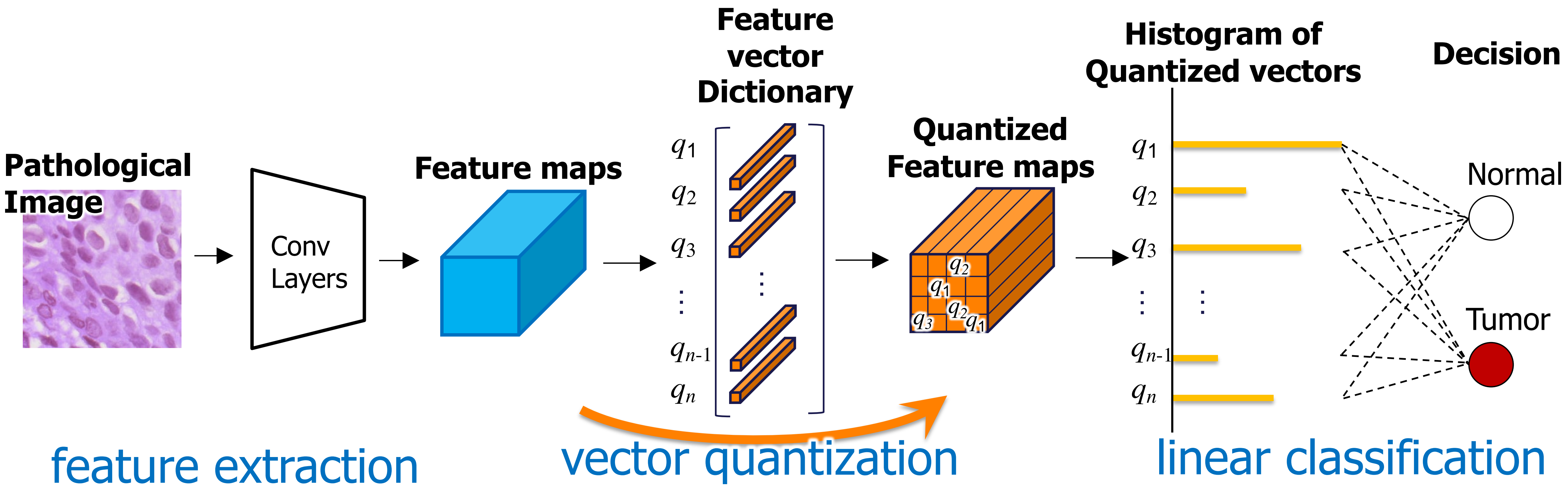


Multi-scale Explainable Feature Learning



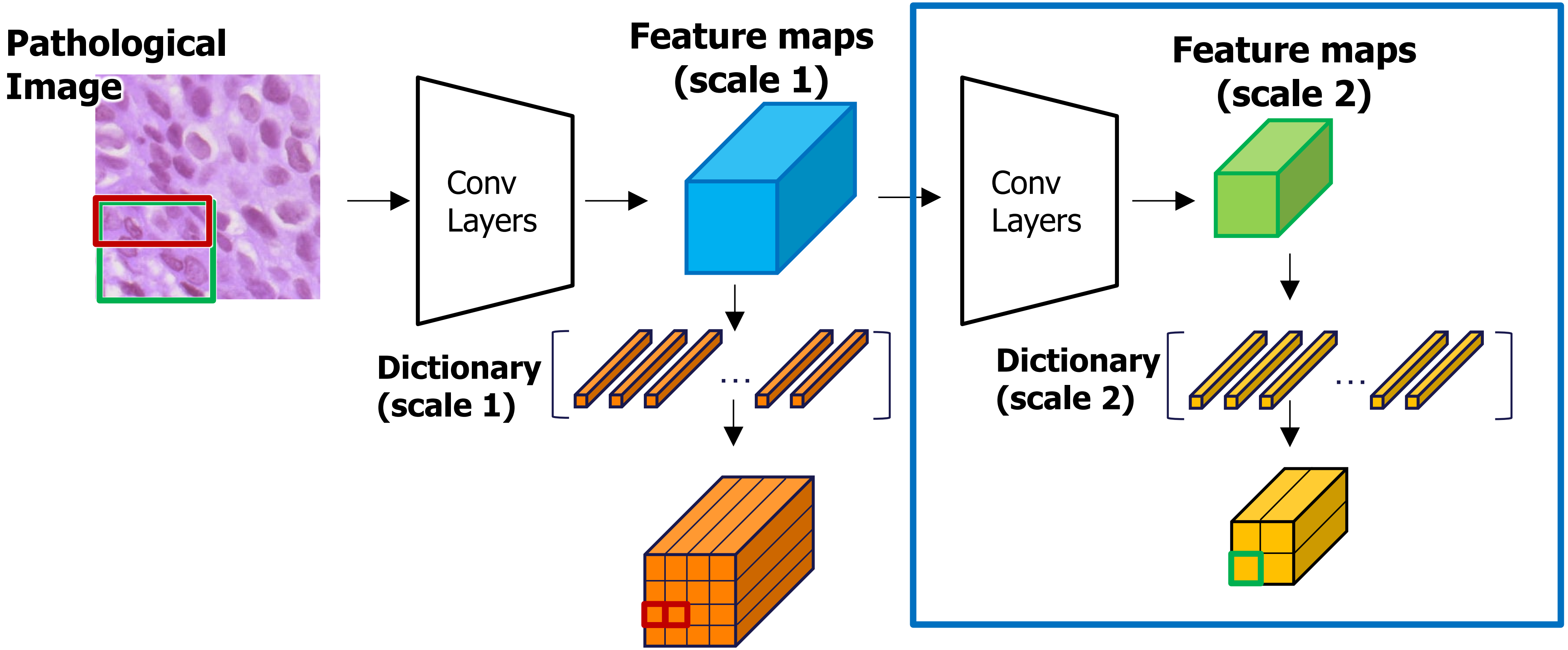
Explainable Classification

- Constructing ***dictionary*** of pathological features from dataset
- Classification is based on the **amount of dictionary items** in an image



Multi-scale Feature Extraction

Multiple dictionaries at different scale allow the network to extract features correctly
Feature vectors in downscaled feature maps cover wider part of original image



Explanations for Basis of Diagnosis

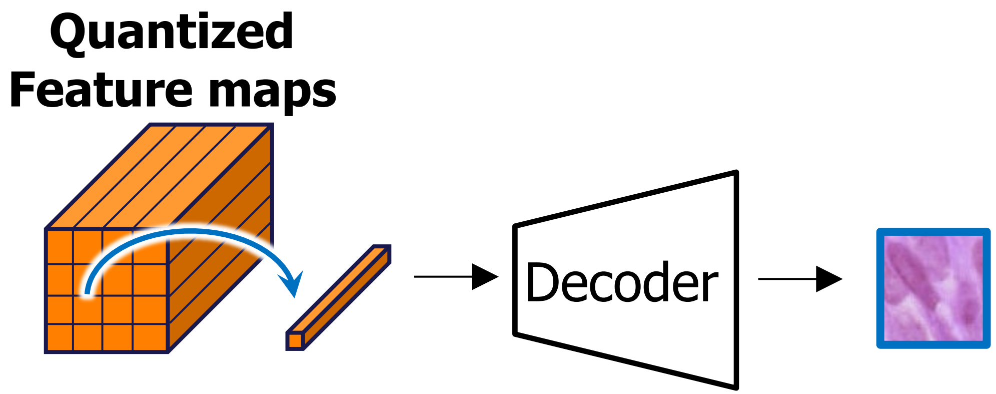
Our method provides two types of explanations

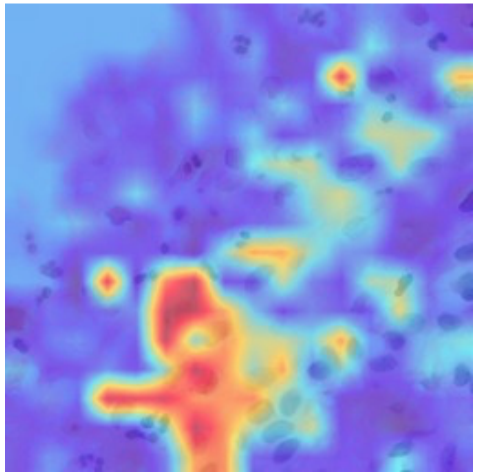
1. Image of quantized features

Users can visually understand which features in the dictionary contributed to decisions

2. Contribution map

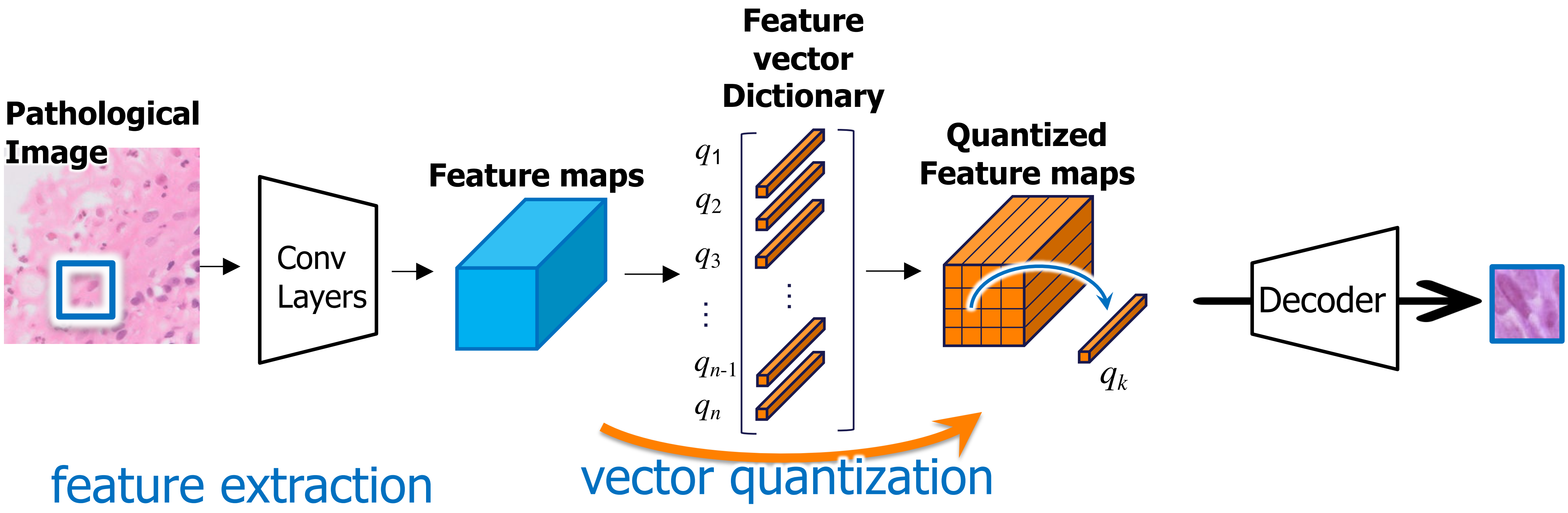
Users can visually understand the part **where** contributed to decisions





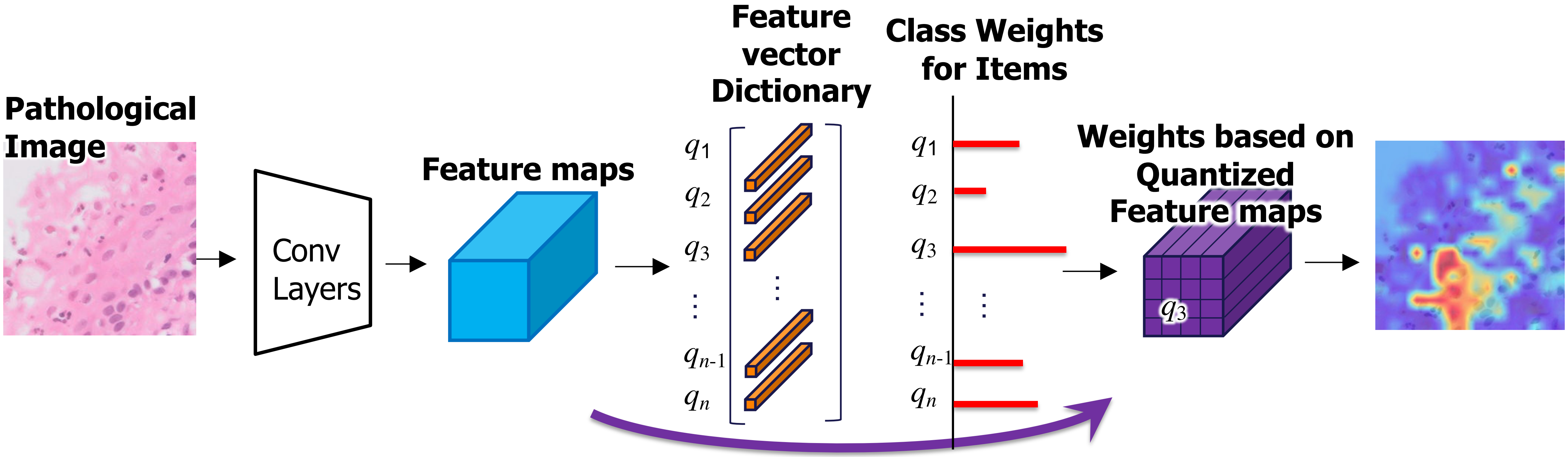
Images of Quantized Features

CNN decoder visualizes items in the feature dictionary from embedded feature space to image space



Contribution Map

Contribution map presents important parts for classification in an image



feature extraction

Mapping weights according to **similarity**

Experimental Setup

Comparison methods

- ✓ Inception V3
- ✓ Prototype based CNN [Uehara+, 2019]
- ✓ Our methods (single scaled feature extraction)

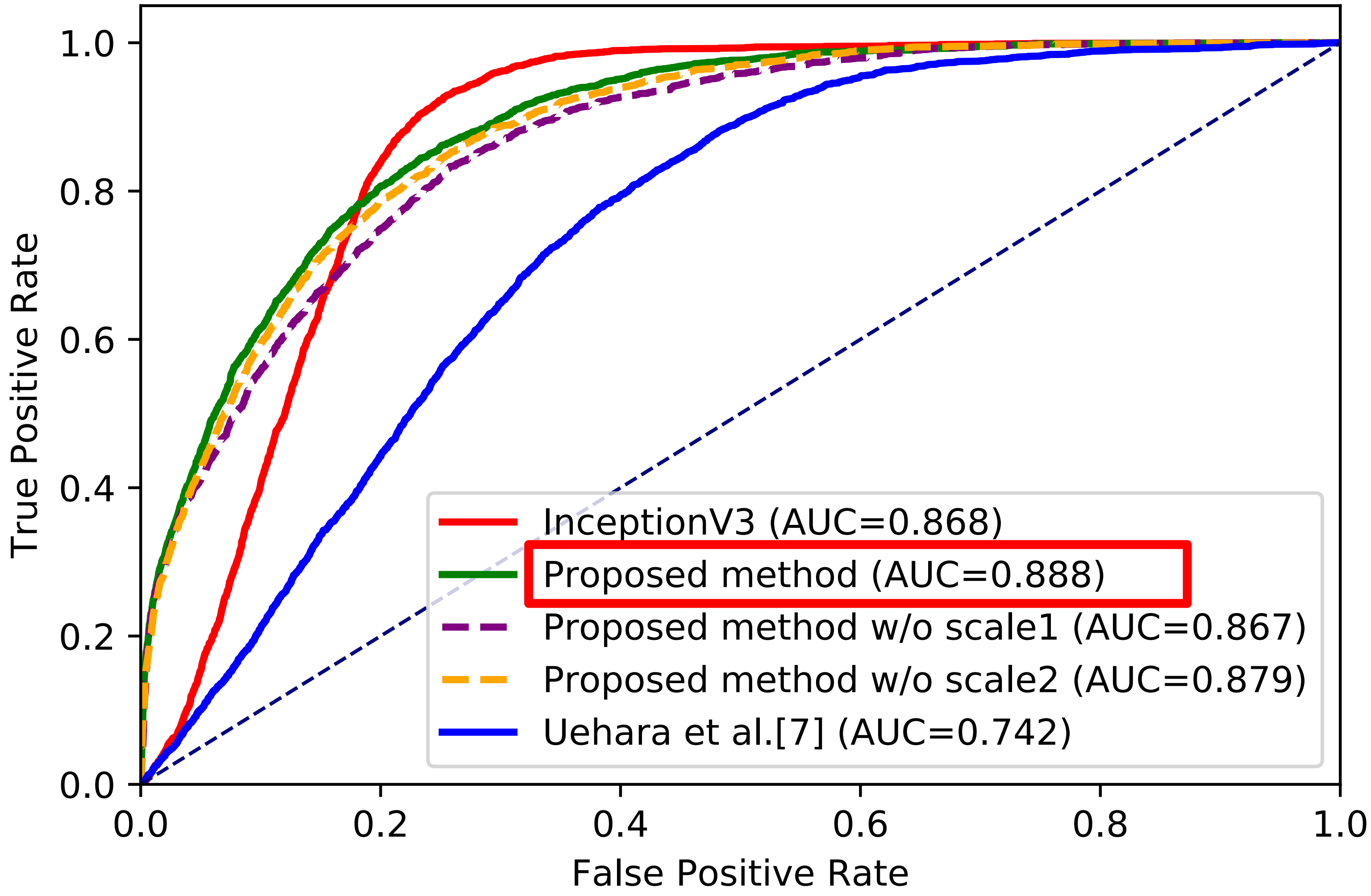
Dataset

Pathological image patches of a uterine cervix
Each image patch has 256x256 pixel
Normal : Train (67,805) Test (15,955)
Abnormal: Train (13,143) Test (2,451)

Classification

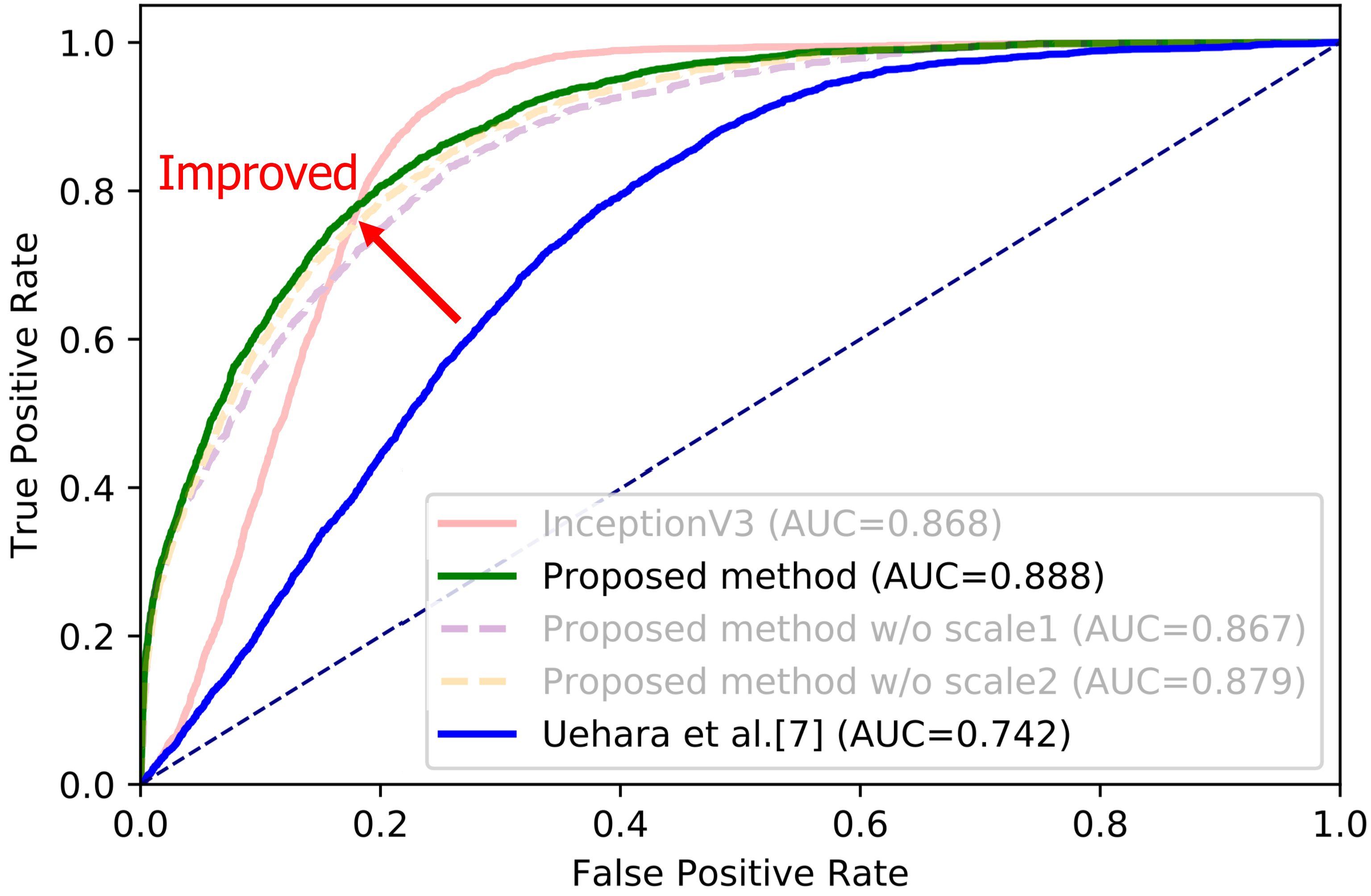
Classification Result

Our method yielded the highest AUROC



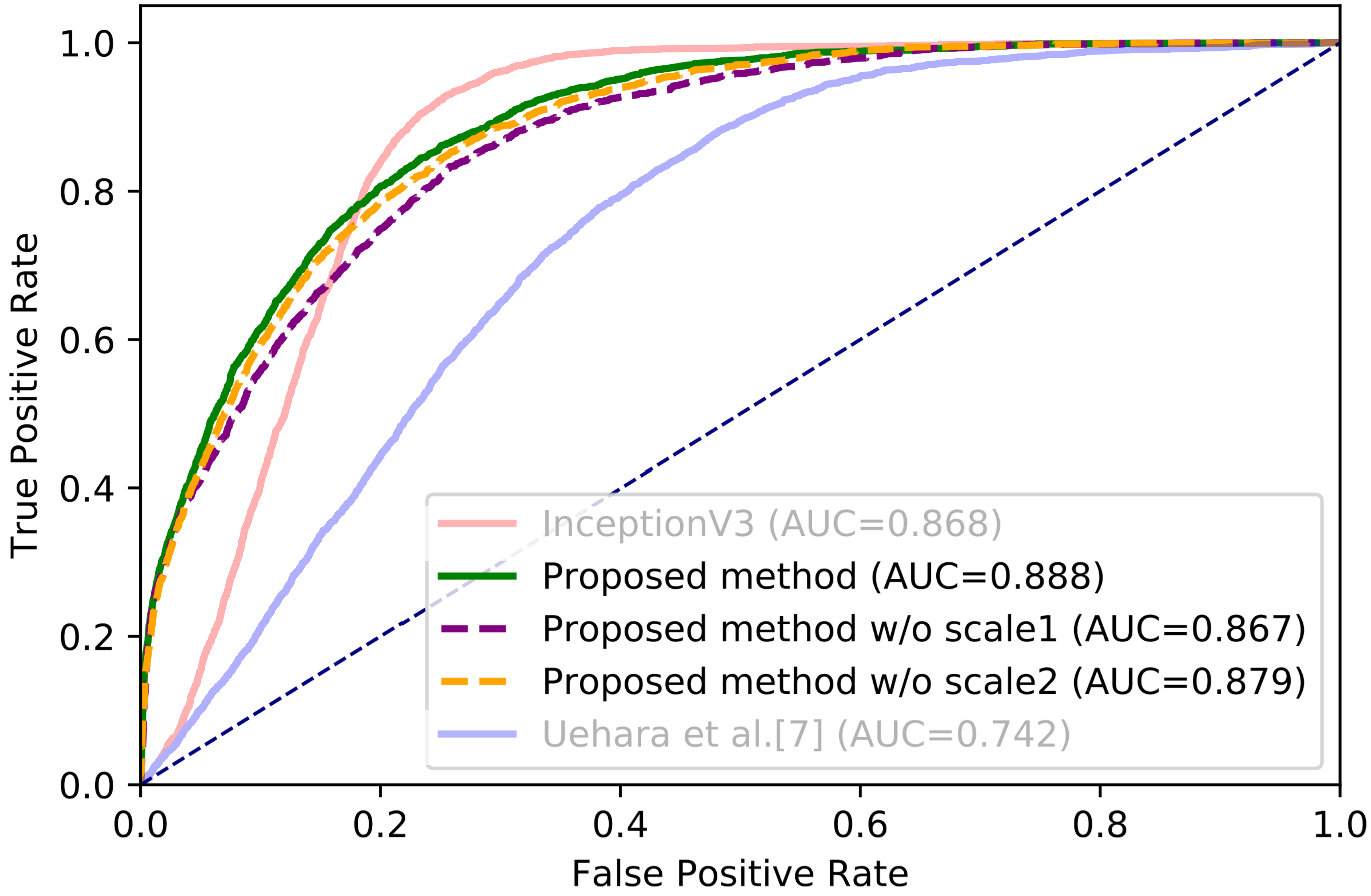
Compared with Explainable CNN

Constructing dictionaries via end-to-end learning makes a good classification

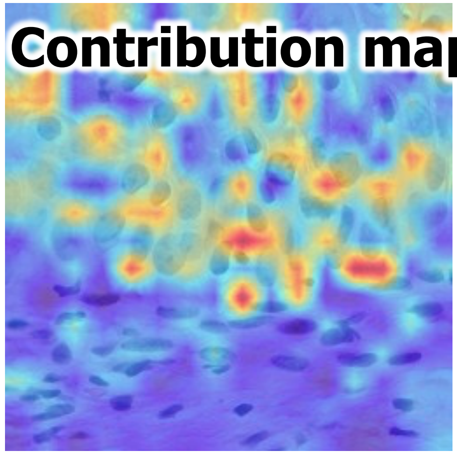


Compared with Single Scale Methods

Combining multi-scale dictionaries shows the best classification

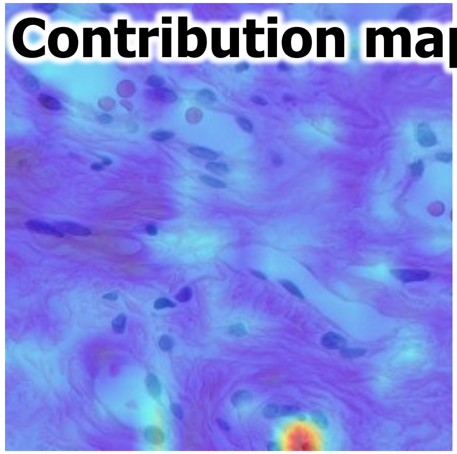


Results of Visual Explanations

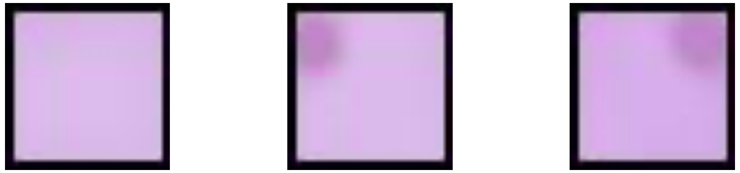


Contributed features

large nuclei



Contributed features



small nucleus

Quality of decoded images
should be improved

Conclusion

We have proposed multi-scale explainable feature learning method to ensure reliable diagnosis for pathological image analysis

✓ **Accurate**
- ➢ Multiple dictionaries for multi-scale features
- ➢ End-to-end dictionary learning

✓ **Easy to interpret its basis of diagnosis**
- ➢ Linear combination of cooccurrence of items in dictionaries
- ➢ Visualize the items as images

Experimental result demonstrated that our method has advantages of **explainability** compared with the conventional **black-box** models