



@ThaoDao2020

Vision-based Assistance for Vocal Fold Identification in Laryngoscopy with Knowledge Distillation

Thao Dao

- *Graduate Student in Information & Communication Technology - Artificial Intelligence*
University of Science - Vietnam National University - Vietnam
- *Otolaryngologist*
Thong Nhat Hospital – Vietnam





Vision-based Assistance for Vocal Fold Identification in Laryngoscopy with Knowledge Distillation



Thao Dao^{1,2,3,4}, Minh-Khoi Pham^{1,2}, Mai-Khiem Tran^{1,2,3}, Bich Tran⁵, Ngoc Van⁶, Chanh Cong⁷, Minh-Triet Tran^{1,2,3}

¹University of Science, Ho Chi Minh City, ²Viet Nam National University, Ho Chi Minh City, ³John von Neumann Institute,

⁴Thong Nhat Hospital, ⁵Cho Ray Hospital, ⁶Vinmec Central Park International Hospital, ⁷A Military Hospital



OUTLINE

1. Introduction
2. Dataset and Evaluate on backbones
3. Implement knowledge distillation
4. Portable laryngoscope with AI smart assistance



I. Introduction

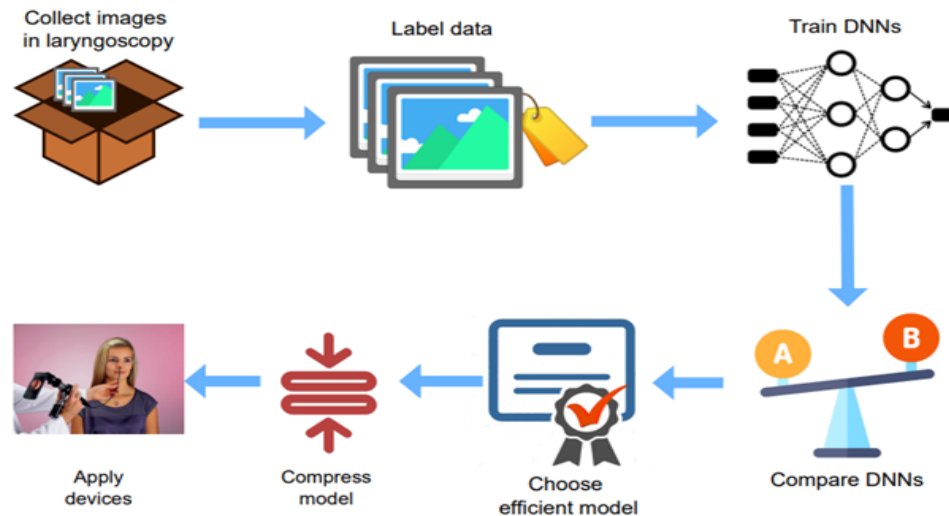
❖ Purposes

- Introduce a **novel dataset** about vocal fold images, along with baseline performance and metrics on a number of backbones.
- Propose a **lightweight network** using **knowledge distillation**.
- Propose a solution of **AI assistance on smartphones** for **laryngoscopy**.



I. Introduction

❖ Overview of our process



Develop a smart vision-based assistance for vocal fold detection and localization



II. Dataset and Evaluate on backbones

❖ Dataset

- Collect 4,624 images of 876 different patients at Department of Otorhinolaryngology, Cho Ray Hospital.
- Filter samples to discard noisy images.
- Two junior ENT doctors independently label images after that revise once again by an expert.



II. Dataset and Evaluate on backbones

❖ Evaluate on backbones

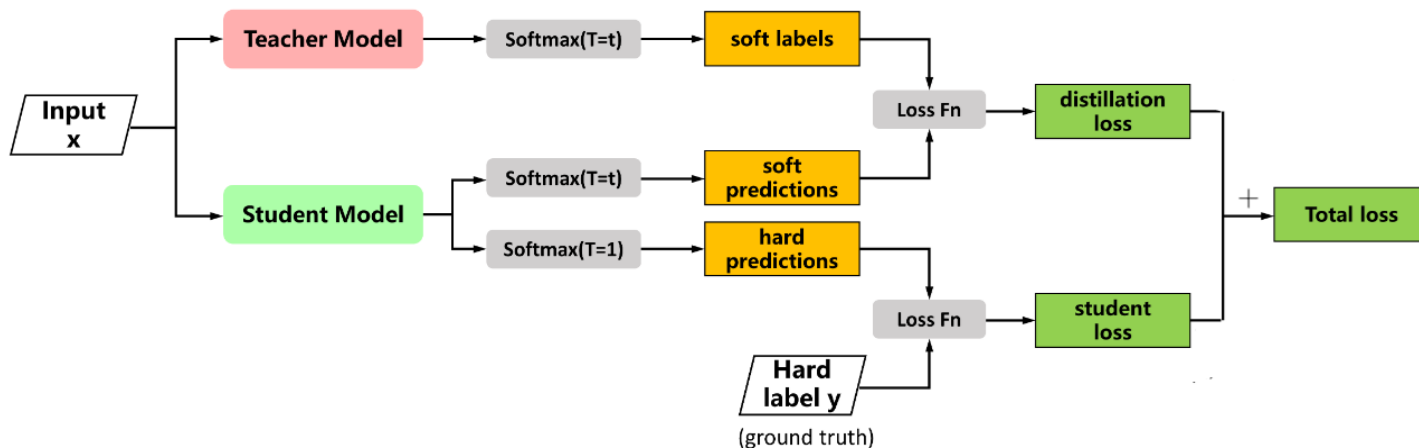
EfficientNetB1 shows extremely **effective** results in assessing images of vocal fold appearance.

| | VGG19 | ResNet50V2 | MobileNetV2 | InceptionV3 | DenseNet201 | Xception | EfficientNetB1 |
|---------------------|-------|------------|-------------|-------------|-------------|----------|----------------|
| Accuracy (%) | 91.8 | 98.5 | 96.1 | 98.3 | 98.2 | 98.2 | 98.7 |
| Recall (%) | | | | | | | |
| Non vocal fold | 94.2 | 98.8 | 96.7 | 98.3 | 98.3 | 98.5 | 99.2 |
| Vocal folds | 88.7 | 98.0 | 95.3 | 98.3 | 98.0 | 97.8 | 98.0 |
| Precision(%) | | | | | | | |
| Non vocal fold | 91.4 | 98.5 | 96.4 | 98.6 | 98.5 | 98.3 | 98.5 |
| Vocal folds | 92.3 | 98.5 | 95.8 | 97.8 | 97.8 | 98.0 | 99.0 |



III. Implement knowledge distillation

❖ Overview Knowledge Distillation



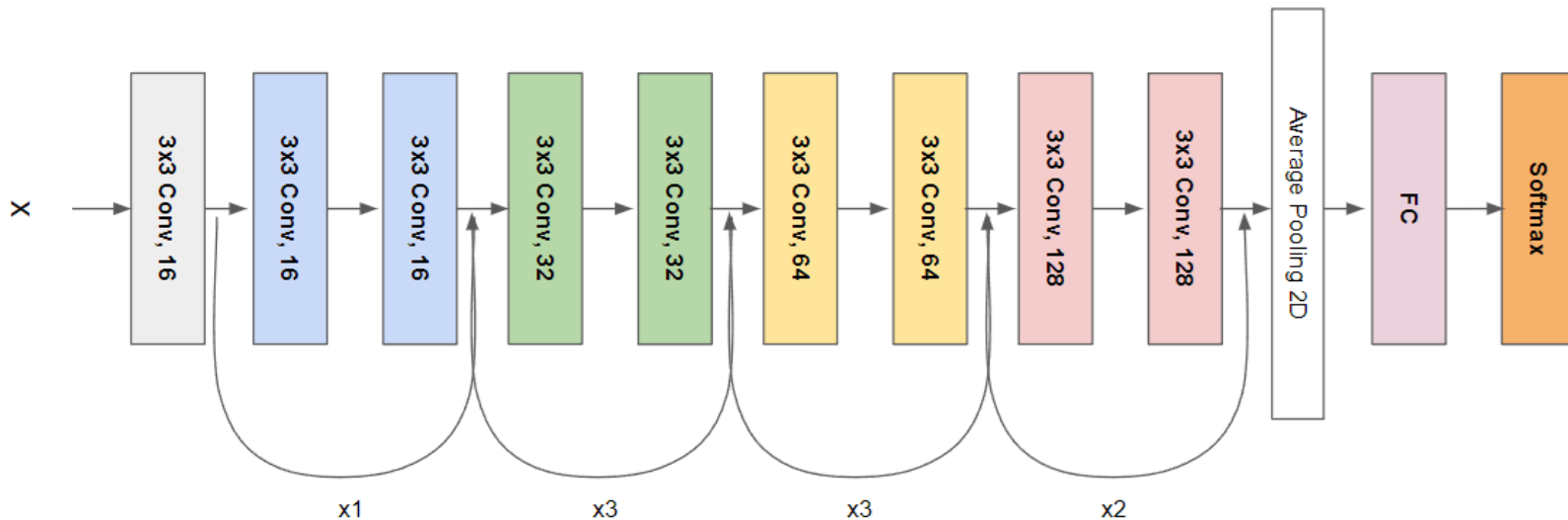
$$\alpha * CE(y_{gt}, y_s) + (1 - \alpha) * KL(y_t, y_s)$$

$$\text{with } KL(y_t, y_s) = F(y_t) \times \log\left(\frac{F(y_t)}{F(y_s)}\right)$$



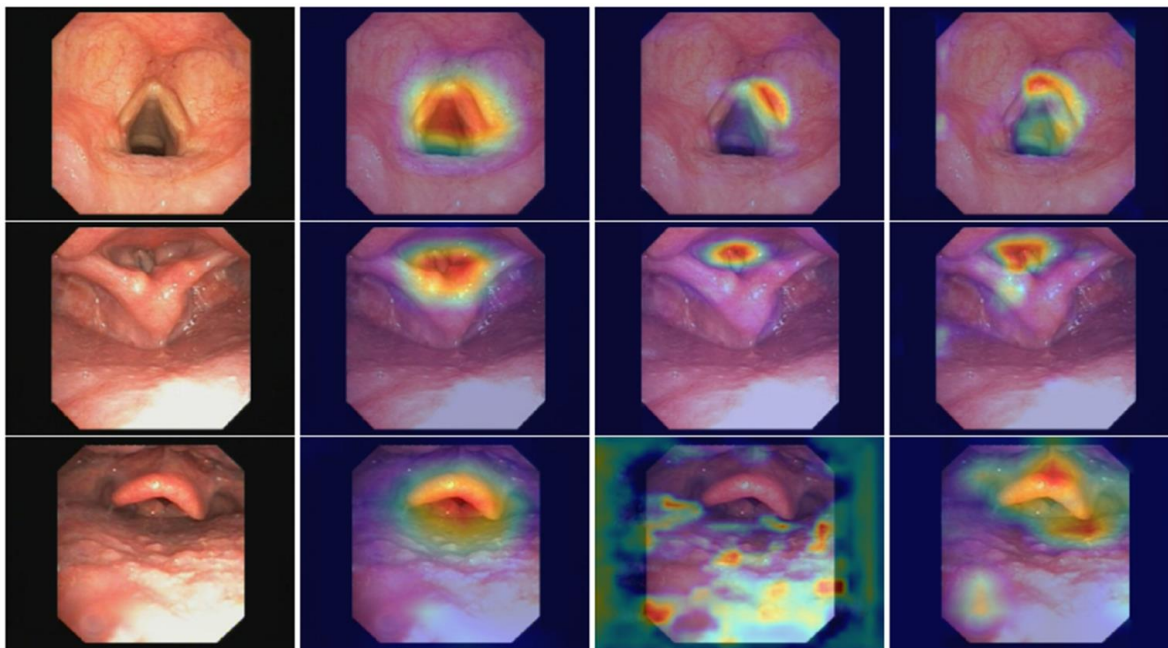
III. Implement knowledge distillation

❖ Architecture of our simple ResNet distilled





III. Implement knowledge distillation



(a) Input image

(b) EfficientNetB1

(c) Simple-ResNet-Scratch

(d) Simple-ResNet-Distilled



III. Implement knowledge distillation

Distillation helps:

- The student's performance to **approximately match the teacher's** one.
- **Faster convergence.**
- **Require minimal computing resource.**

| Methods | Accuracy | No. parameters |
|-------------------------|----------|----------------|
| EfficientNet-B1 | 98.7% | 6.7M |
| MobileNetV2 | 96.1% | 2.4M |
| Simple-ResNet-Scratch | 96.7% | 0.8M |
| Simple-ResNet-Distilled | 98.4% | 0.8M |



IV. Portable laryngoscope with AI smart assistance

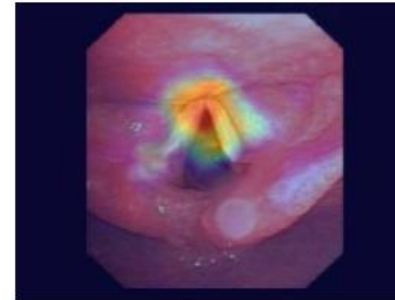
- A system apply our distilled deep learning model to portable laryngoscopy devices.
- This system includes:
 - ✓ Mobile application that integrates our distilled model.
 - ✓ Smartphone.
 - ✓ Endoscope adaptor.
 - ✓ Light source.
 - ✓ Laryngeal endoscope.



IV. Portable laryngoscope with AI smart assistance



Images from laryngoscope connected to smartphone



Vocal folds located in image in realtime



Vision-based Assistance for Vocal Fold Identification in Laryngoscopy with Knowledge Distillation

Thao Dao^{1,2,3,4}, Minh-Khoi Pham^{1,2}, Mai-Khiem Tran^{1,2,3}, Bich Tran⁵, Ngoc Van⁶, Chanh Cong⁷, Minh-Triet Tran^{1,2,3}

¹University of Science, Ho Chi Minh City, ²Viet Nam National University, Ho Chi Minh City, ³John von Neumann Institute, ⁴Thong Nhat Hospital, ⁵Cho Ray Hospital, ⁶Vinmec Central Park International Hospital, ⁷A Military Hospital.

In our research:

- Introduce a novel dataset about vocal fold images, along with baseline performance and metrics on a number of backbones.
- Propose a lightweight network using knowledge distillation and our student model achieves approximately the same accuracy.
- Propose a solution of AI assistance on smartphones.

Image

Drop Image Here
or
Click to Upload

Output

Flag

Model

Which model?

VGG19 ResNet50V2

MobileNetV2 InceptionV3

DenseNet201 Xception

EfficientNetB1 ResNet-Distilled

Clear Submit

Examples

Pages 12



Vision-based Assistance for Vocal Fold Identification in Laryngoscopy with Knowledge Distillation

Thao Dao^{1,2,3,4}, Minh-Khoi Pham^{1,2}, Mai-Khiem Tran^{1,2,3}, Bich Tran⁵, Ngoc Van⁶, Chanh Cong⁷, Minh-Triet Tran^{1,2,3}

¹University of Science, Ho Chi Minh City, ²Viet Nam National University, Ho Chi Minh City, ³John von Neumann Institute, ⁴Thong Nhat Hospital, ⁵Cho Ray Hospital, ⁶Vinmec Central Park International Hospital, ⁷A Military Hospital.

In our research:

- Introduce a novel dataset about vocal fold images, along with baseline performance and metrics on a number of backbones.
- Propose a lightweight network using knowledge distillation and our student model achieves approximately the same accuracy.
- Propose a solution of AI assistance on smartphones.

Image

Output

Vocal Folds

Vocal Folds 100%

No Vocal Folds 0%

Flag

Model

Which model?

VGG19 ResNet50V2

MobileNetV2 InceptionV3

DenseNet201 Xception

EfficientNetB1 ResNet-Distilled

Clear Submit

Examples

Pages 12




Summary

- A **novel dataset** about the laryngoscopy images.
- A **suitable model** for applying **knowledge distillation**.
- **AI assistance on smartphones** to create a portable laryngoscope system.

The logo for MEDINFO 23, featuring the word 'MED' in a colorful, mosaic-like font, followed by 'INFO' in a bold, white sans-serif font, and '23' in a large, white, thin-lined font.

MEDINFO23

8 - 12 JULY 2023 | SYDNEY, AUSTRALIA

A person wearing a white lab coat and white gloves is shown in a laboratory setting, carefully handling a small blue-capped vial. The background is slightly blurred, showing other laboratory equipment and a rack of similar vials. The overall scene is brightly lit, emphasizing the clinical and scientific nature of the event.

Thanks for your attention