netmarble

ICASSP2020 Barcelona

Robust End-to-end Keyword Spotting and Voice Command Recognition for Mobile Game

Hu Xu*, Youshin Lim*, Shounan An, Hyemin Cho, Yoonseok Hong, Insoo Oh

Magellan Division, AI Center, Netmarble

6th May 2020

*Both authors contributed equally to this work

Table of contents

netmarble

Motivation

Why do we need a voice UI for mobile game?

Deep learning(DL) based speech recognition is ready for mobile game?

DL is truly a technological enabler, but need to be more developed for **on-device speech recognition for mobile game**.

netmarble

Keyword Spotting(KWS)

Problem definition: we integrate KWS into mobile game, which used as a voice UX for gamers when their hands are busy.

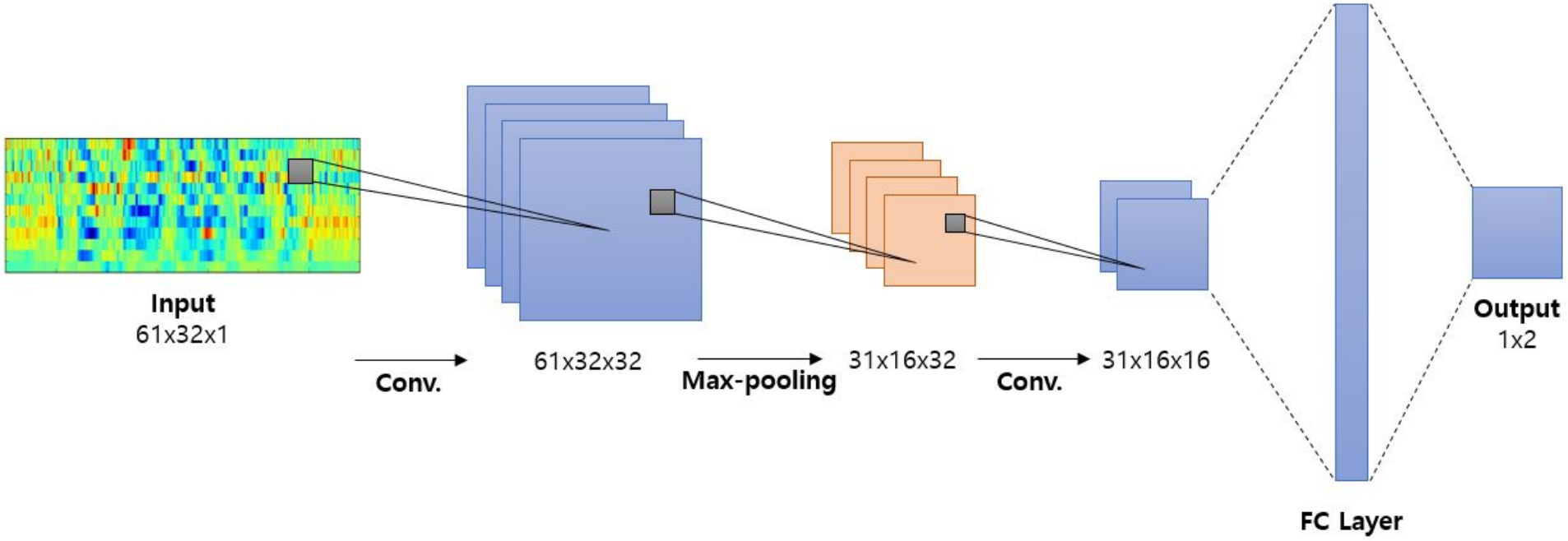
Game: A3:STILL ALIVE
Company: Netmarble & IDEA Games
Genre: Massively Multiplayer Online Role-Playing Game(MMORPG)
Launch date: 12th March 2020

Before KWS

After KWS

KWS: challenge

*Challenge: mobile phone should **not to be overheated** with KWS, therefore a computationally very light deep neural network is mandatory.*



KWS SDK

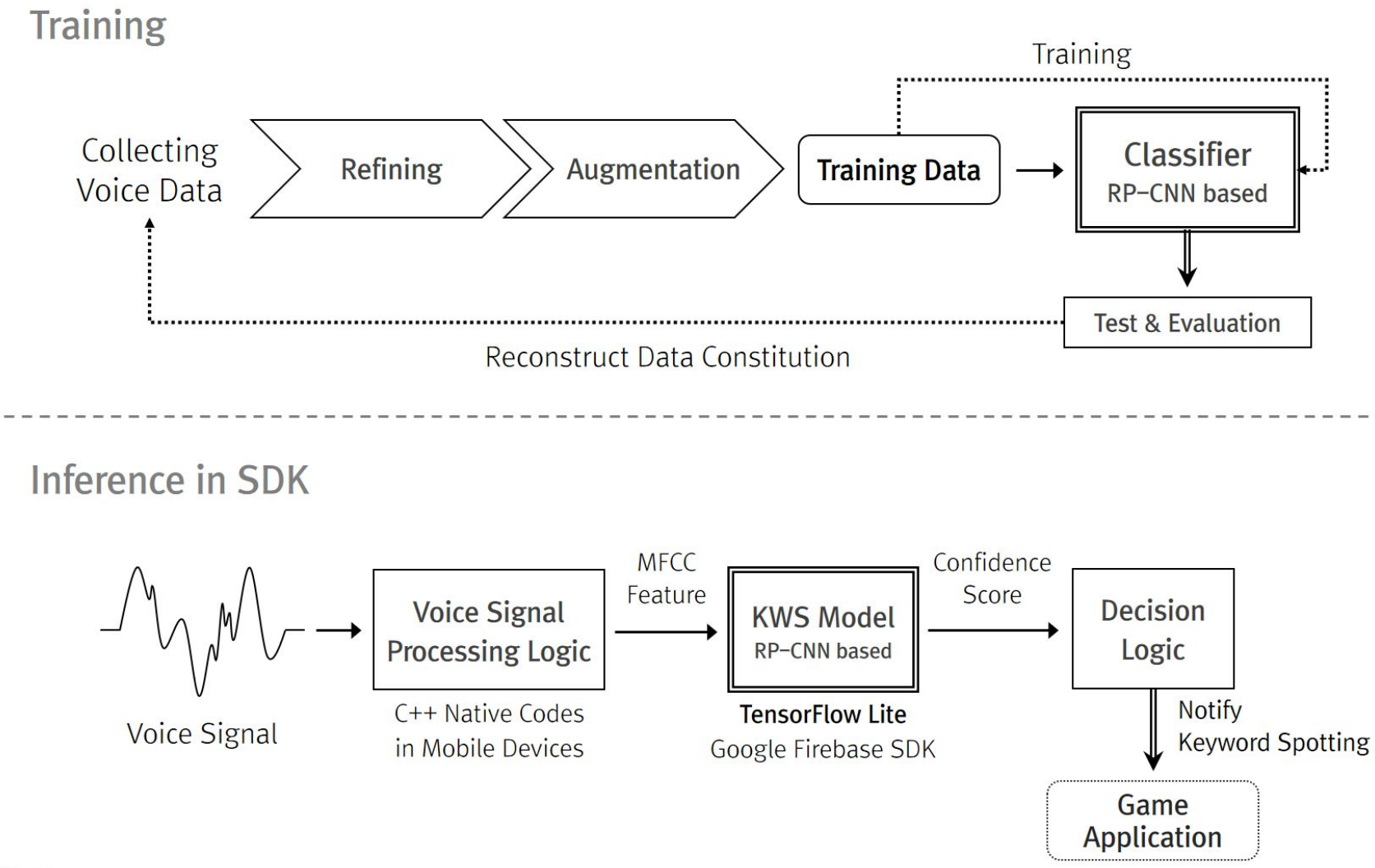


Figure from "Robust keyword spotting via recycle-pooling for mobile game", Interspeech 2019'

KWS: how to improve accuracy?

Datasets

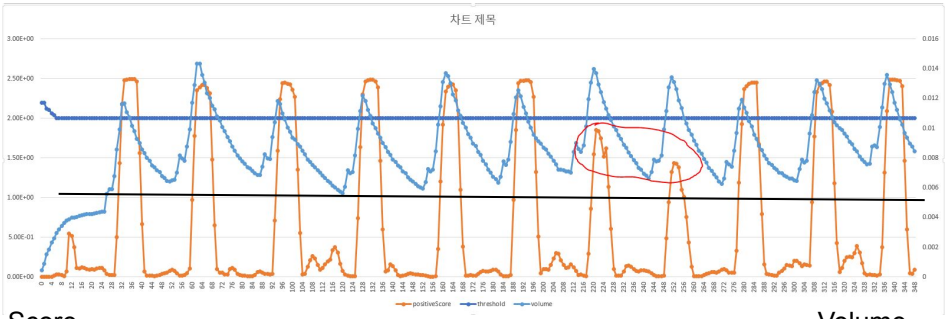
Keyword: “Monica”
Training data: 180 people
Data acquisition: android and iPhone in various environments(office, café, street etc.)
Data augmentation: various volume, mixing with background noise etc.





Improve recall

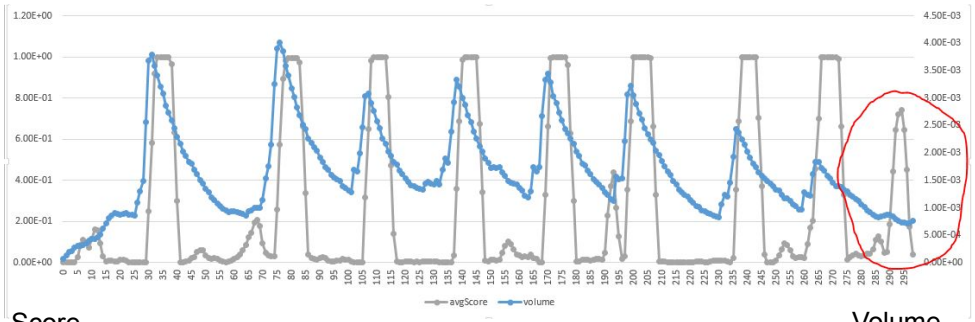
Adaptive thresholding with average of inference score and summation of absolute difference of continuous frames.



Score Volume

Improve precision

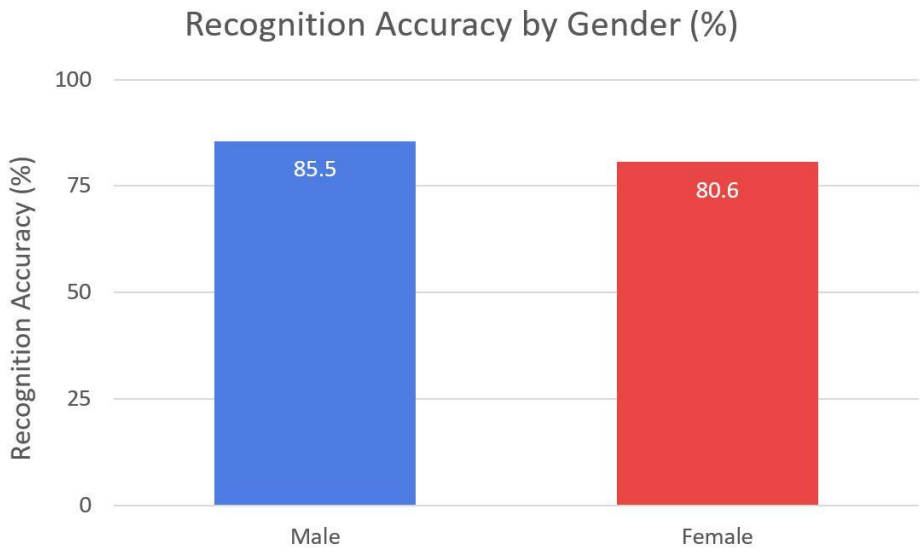
We consider volume of coming speech signal as well to reject outliers from various noise.

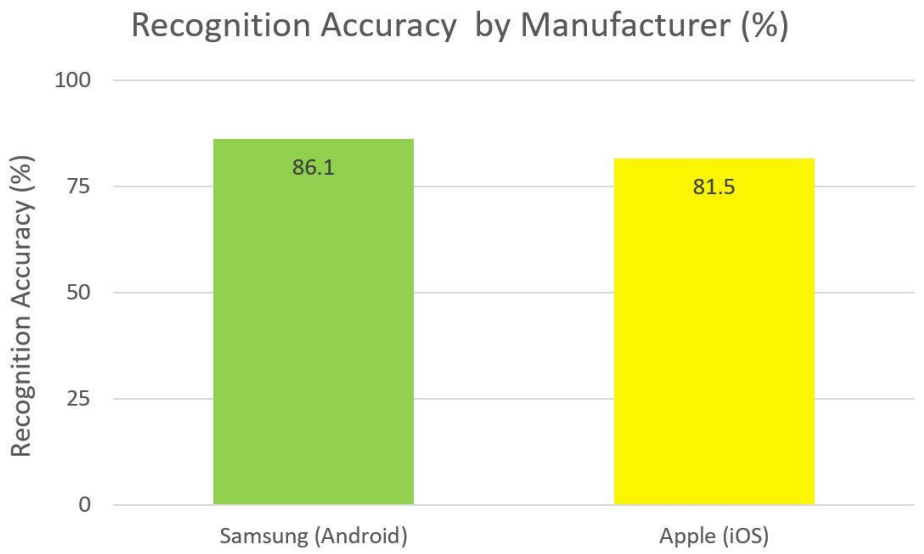


Score Volume

netmarble

KWS: accuracy





Test: 27 people, test datasets is 15% of training counterpart.

netmarble

KWS: performance

Android deployment: INT8 quantization with

iOS deployment: FP32 with

AOS	Performance
KWS weight file	60 KB
Memory usage	5~6 MB
CPU usage	< 1%
Device	Samsung Galaxy S8

iOS	Performance
KWS weight file	224 KB
Memory usage	6~7 MB
CPU usage	1%
Device	iPhone 8

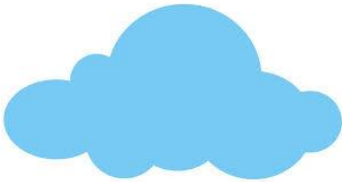
Demonstration

Game Command Recognition(GCR)

Motivation: we need an immediate responding speech recognition for intensive mobile game users, therefore on-device GCR is necessary.

"Start Auto Battle"

cloud based
speech recognition



session time: 15 seconds

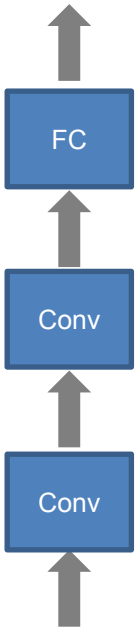
too late...

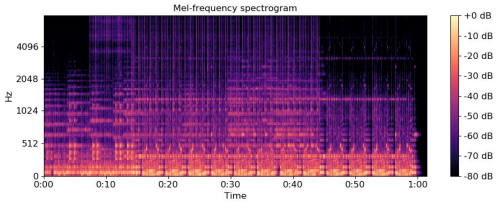
netmarble

GCR: architecture

DNN model: we choose Transformer as our speech recognition module for its high accuracy and relatively easier mobile deployment.

positional encoding





feature: melspectograms

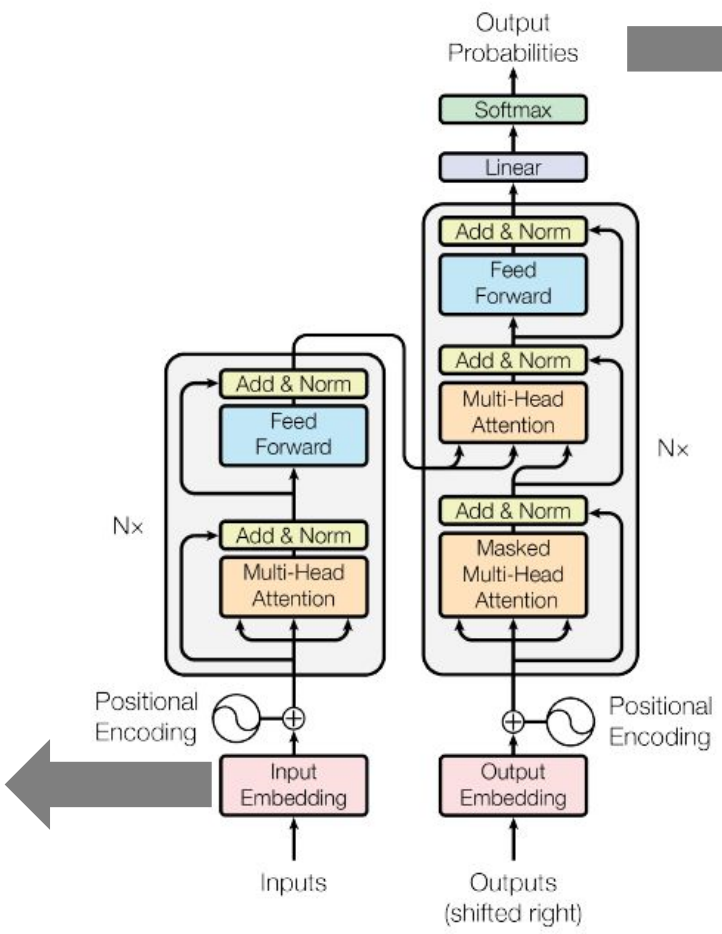
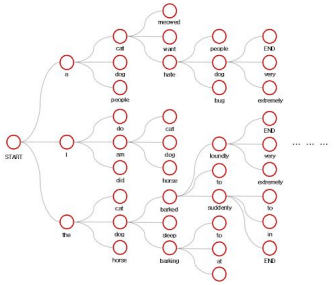


Figure from “Attention is All You Need”, NeurIps 2017’

Post-processing:

- Beam search



- Measure cos similarity within game command lists and return the highest command as final result.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

netmarble

GCR: data

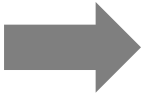
Korean	datasets size	# people	remark
AIHub	1000 hours	< 1000	publicly open
NIKL	136 hours	40	publicly open
from TTS	5.1 hours	62	generated inhouse
Magellan datasets	1.3 hours	22	collected inhouse

AIHub : http://www.aihub.or.kr/aidata/105/download
NIKL : https://ithub.korean.go.kr/user/total/referenceManager.do
from TTS: voice generated from Text To Speech(TTS) engine by Netmarble, Google and Kakao

GCR: accuracy & performance

Training framework with ESPnet*

PYTORCH



Mobile deployment with

TensorFlow Lite

	RNN-T	Transformer
Weight file	45.3 MB	76 MB
Accuracy	84.3%	92.8%

Test datasets: AIHub, NIKL, from TTS, Magellan datasets, about 10% of training counterpart.

reference: “ESPnet: End-to-End Speech Processing Toolkit”, Interspeech 2018’

GCR: INT8 quantization with PyTorch v1.4

PyTorch v1.4	current limitation for our Transformer model
Quantization aware training	supports FakeQuantization for CPU only
Static quantization	sin/cos operation of positional encoding in Transformer is not supported yet
Dynamic quantization	no processing time reduction for its on-line calibration step for each data batch

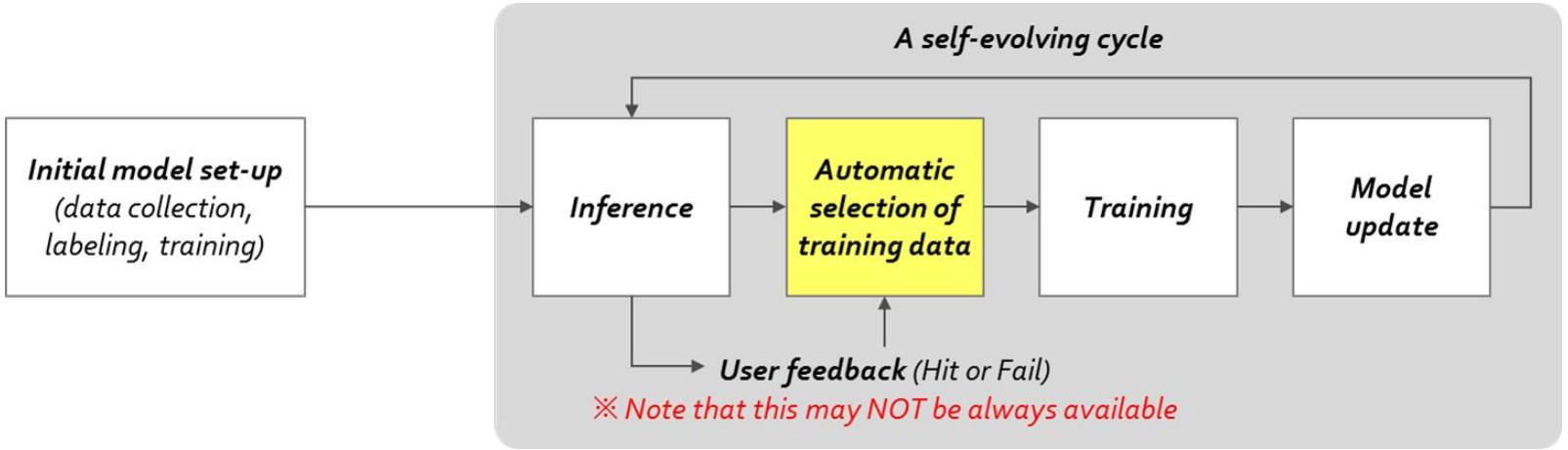
Demonstration

netmarble

Future work

- ***We will further optimize our transformer model and GCR inference module (e.g. INT8 quantization) to improve the performance on mobile device, which is necessary steps to integrate GCR into A3: STILL ALIVE!***

- ***Online incremental learning for individual users***



Thank you

- *We would like to thank members from Netmarble and IDEA Games who donated your valuable voice.*
- *Especially thanks to **Hanwook Lee, Gwangmin Hong** from IDEA Games and **Daegeun Choe** from Netmarble to produce the domentration videos.*