**EFFICIENT KEYWORD SPOTTING USING DILATED CONVOLUTIONS AND GATING**

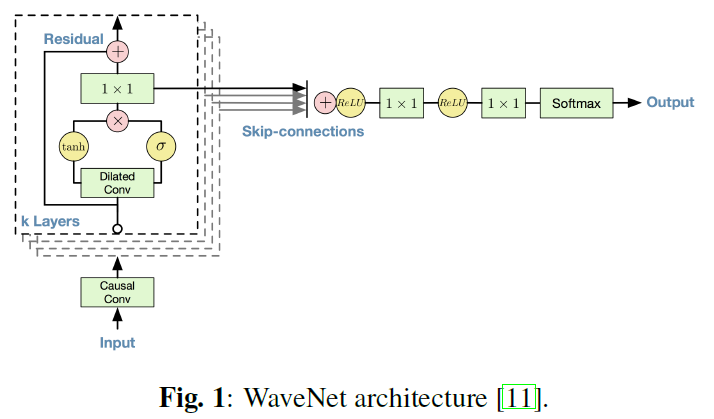
1. INTRODUCTION

Keyword spotting (KWS) aims at detecting a pre-defined keyword or set of keywords in a continuous stream of audio. In this work we focus on end-to-end stateless temporal modeling which can take advantage of a large context while limiting computation and avoiding saturation issues.

2. MODEL IMPLEMENTATION

2.1. System description

2.2. Neural network architecture

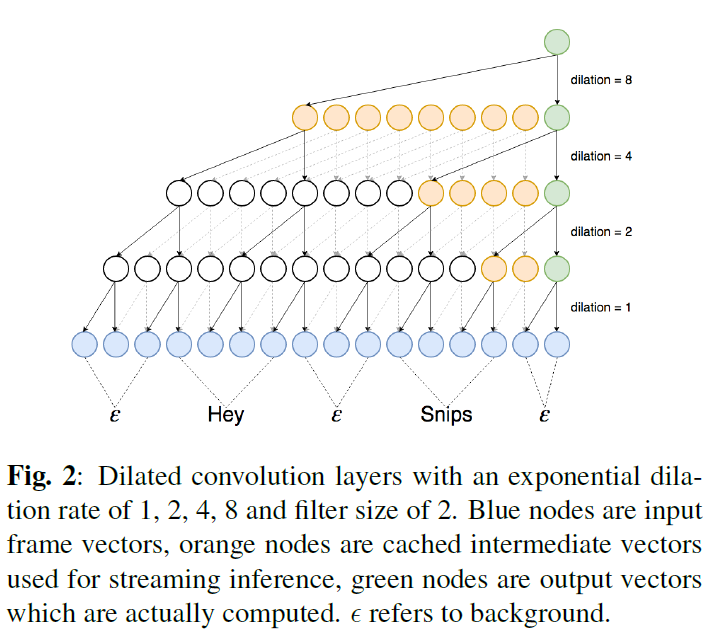


2.2.1. Dilated causal convolutions

Standard convolutional networks cannot capture long temporal patterns with reasonably small models due to the increase in computational cost yielded by larger receptive fields. Dilated convolutions skip some input values so that the convolution kernel is applied over a larger area than its own.

2.2.2. Gated activations and residual connections

Gated activations units – a combination of tanh and sigmoid activations controlling the propagation of information to the next layer – prove to efficiently model audio signals.



2.3. Streaming inference

In addition to reducing the model size, dilated convolutions allow the network to run in a streaming fashion during inference, drastically reducing the computational cost.

2.4. End-of-keyword labeling

3. EXPERIMENTS

3.1. Open dataset

3.2. Experimental setup

3.3. Results

4. CONCLUSION