

语音信号降噪研究方法介绍

海狸谷 音频降噪小组

徐高鹏

C 目录 ontents



1 传统语音降噪方法介绍

2 常规神经网络语音降噪介绍

3 GAN简介

4 GAN在语音增强领域的应用

传统的语音降噪算法

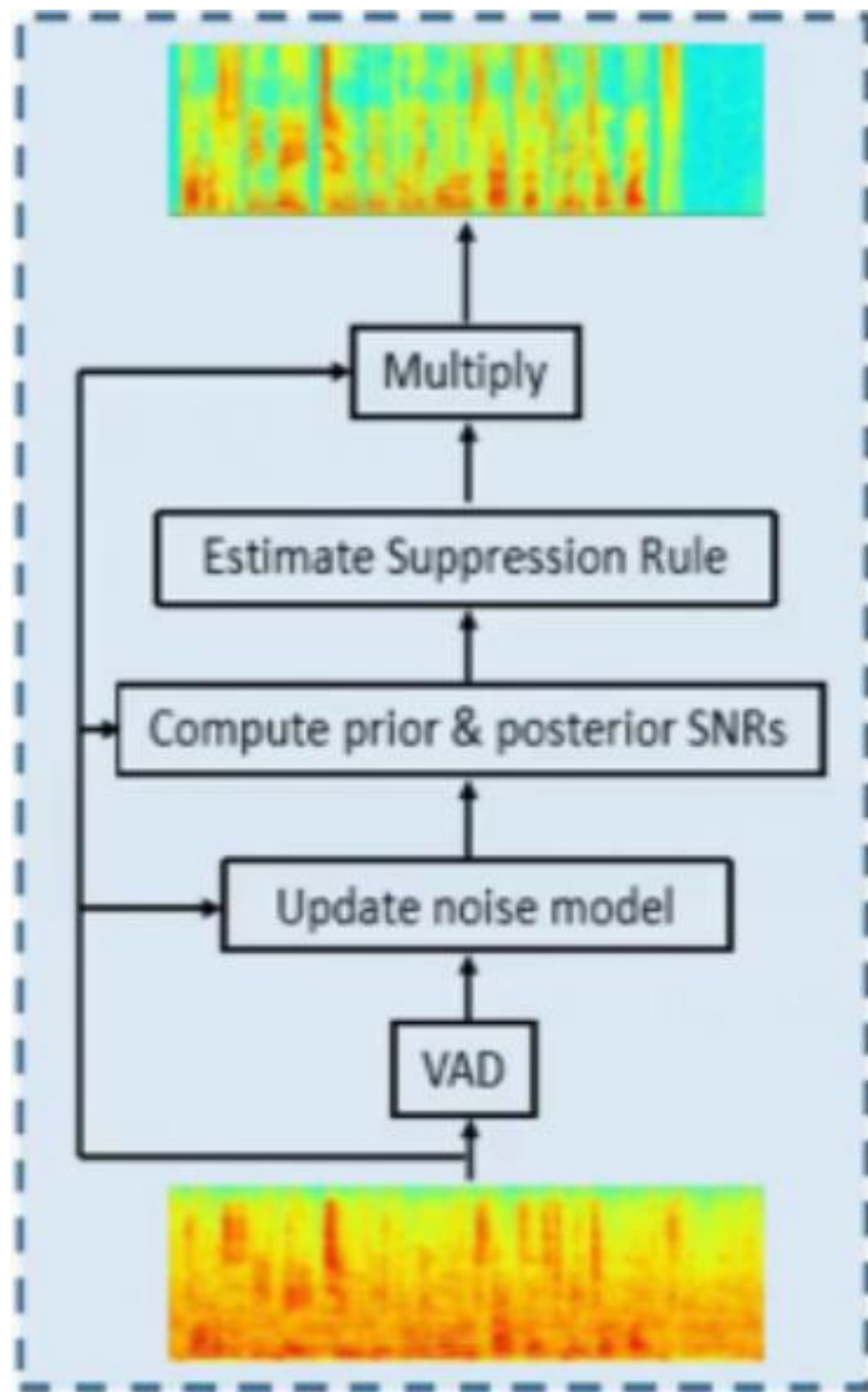
时域 $y(t) = S(t) + n(t)$

频域 $Y_k = S_k + N_K$

方法		设计准则	增益函数(suppression rule)
谱减法		$ S_k = Y_k - N_k $	$G_k = \sqrt{\frac{\xi_k}{\xi_k + 1}}$
维纳滤波		$\arg \min_w E\{ w^H y - s ^2\}$	$G_k = \frac{\xi_k}{\xi_k + 1}$
统计模型 (最大似然, 贝叶斯, MMSE)	Log-MMSE	$\arg \min E\{ \log \hat{S}_k - \log S_k ^2\}$	$G_k = \frac{\xi_k}{\xi_k + 1} \exp\{\frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt\}$

判决引导法 $\xi_k = \alpha \frac{|S_k(t-1)|^2}{\lambda_d(t-1)} + (1 - \alpha) \max[\gamma_k(t) - 1, 0]$ (历史先验SNR和当前先验SNR的加权平均)

传统的语音降噪算法



传统的语音降噪算法的实现步骤:

- 1.分帧，加窗，FFT
- 2.噪声功率谱的估计
- 3.计算对应得先验和后验SNR，利用判决引导法更新先验SNR
- 4.根据对应的增益函数计算增益值
- 5.增益值乘以带噪声的STFT，并IFFT，得到降噪后的时域信号

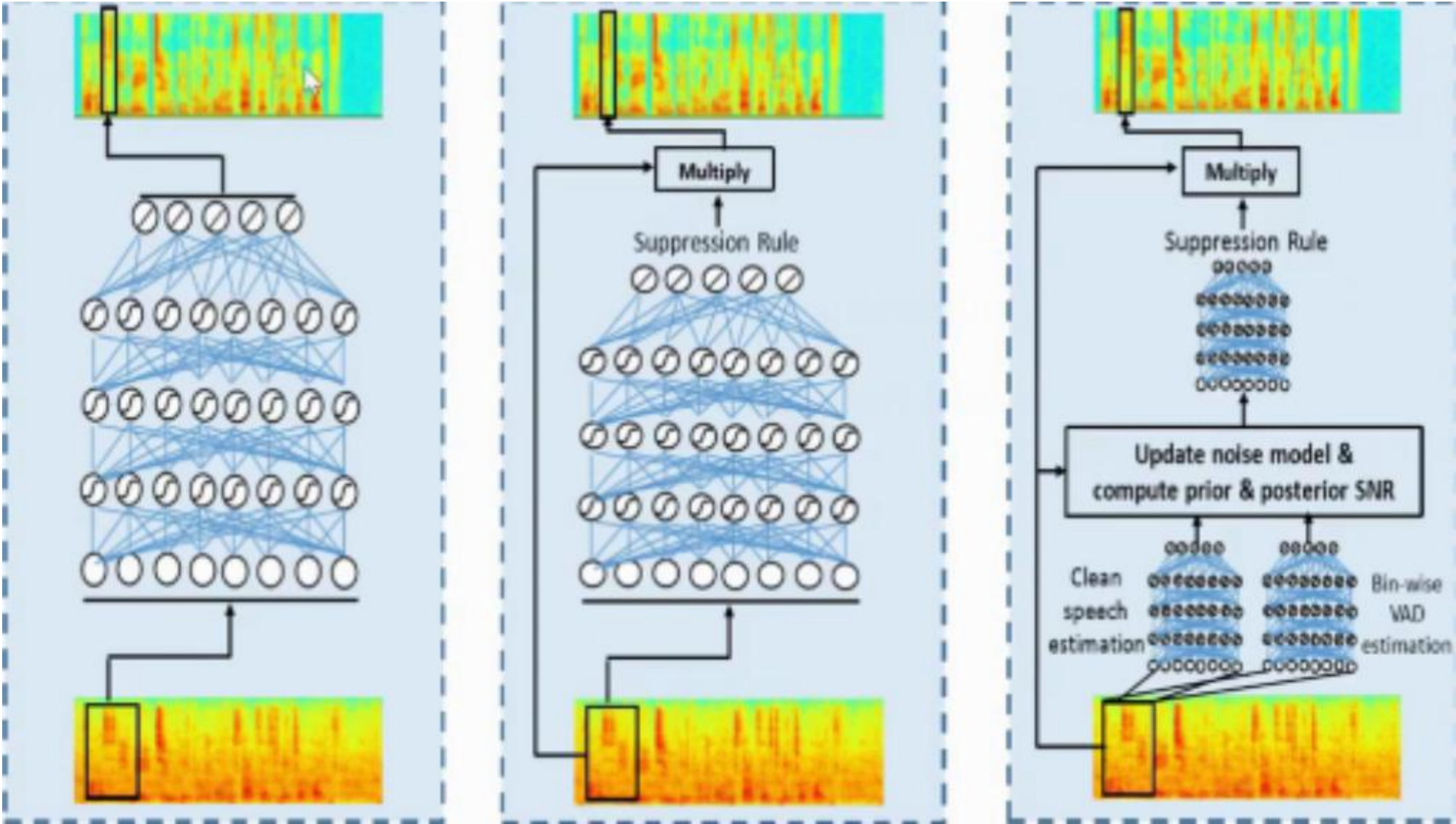
基于神经网络的语音降噪算法

Ideal Binary Mask: 0 or 1

Soft Mask : $G_k = \frac{\xi_k}{\xi_{k+1}}$

Method	input	Ideal_output
分类	每个子带的特征	IBM
回归	当前帧与前后若干帧联合的带噪语音log谱	Soft mask
回归	当前帧与前后若干帧联合的带噪语音log谱	Clean_feature

基于神经网络的语音降噪算法



Lvan Tashev,2017
泛化能力更强

GAN简介

生成对抗网络（Generative Adversarial Network, GAN）最早由Ian Goodfellow 在2014年提出，是目前深度学习领域最具潜力的研究成果之一，它的核心思想是：同时训练两个相互协作，同时又相互竞争的深度神经网络（一个为生成器，一个为判别器）目前已经在图像领域取得很大的成果，但在语音方面，目前应用较少，本次介绍两个关于GAN在语音方面的应用。

GAN简介

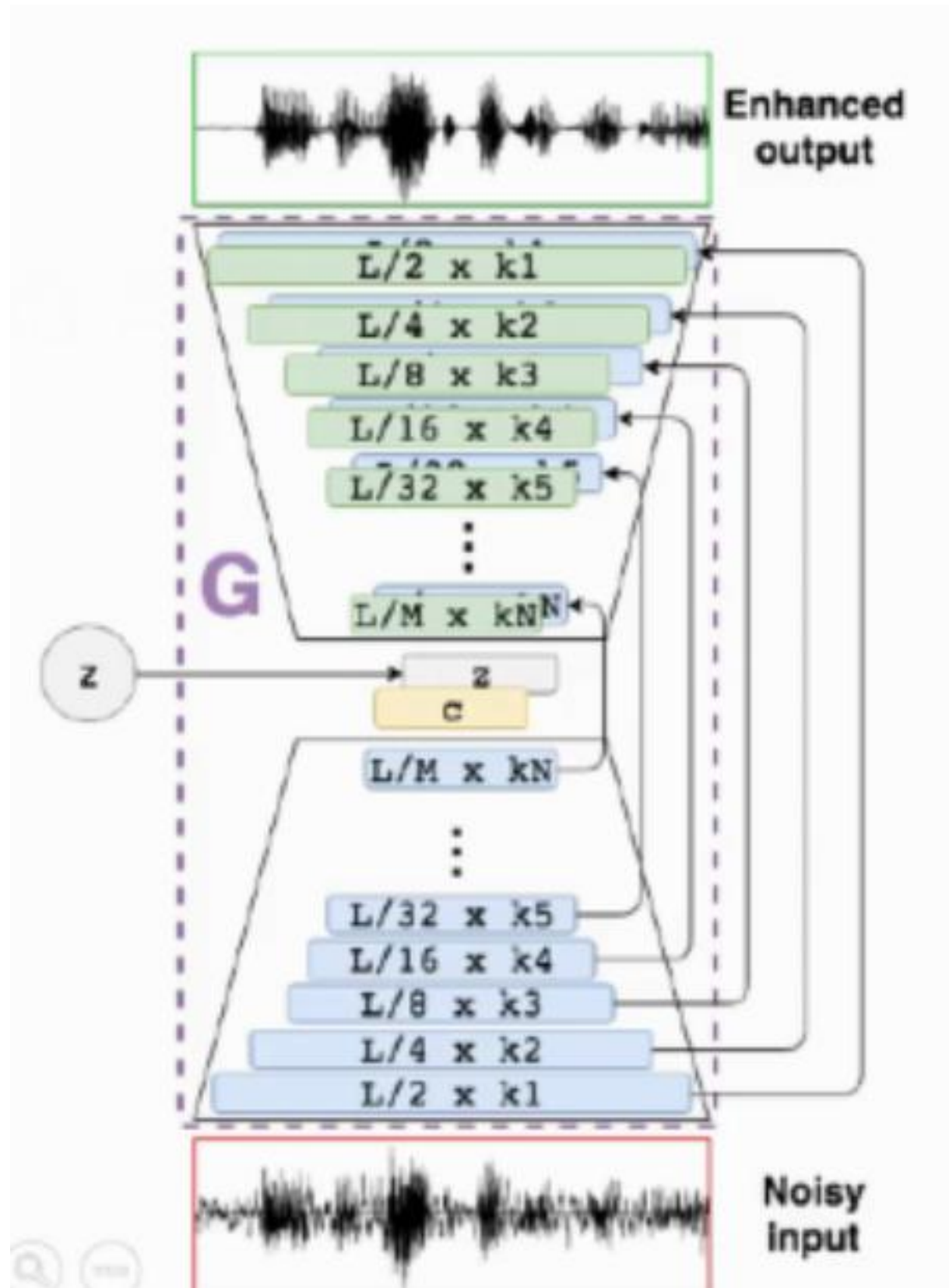


菜鸟画家：生成器

菜鸟鉴赏家：判别器

1. 新手画家画出一幅蒙娜丽莎画作
2. 新手鉴赏家通过和真实画作对比，进行判断。并给出反馈意见
3. 根据鉴赏家的建议，改进工艺，进行画作，然后再给鉴赏家鉴赏
4. 重复整个流程，画家技术越来越好，同时鉴赏家的鉴赏能力也不断提高，直到鉴赏家再区分不出真假

基于GAN语音降噪算法



SEGAN: Speech Enhancement Generative Adversarial Network

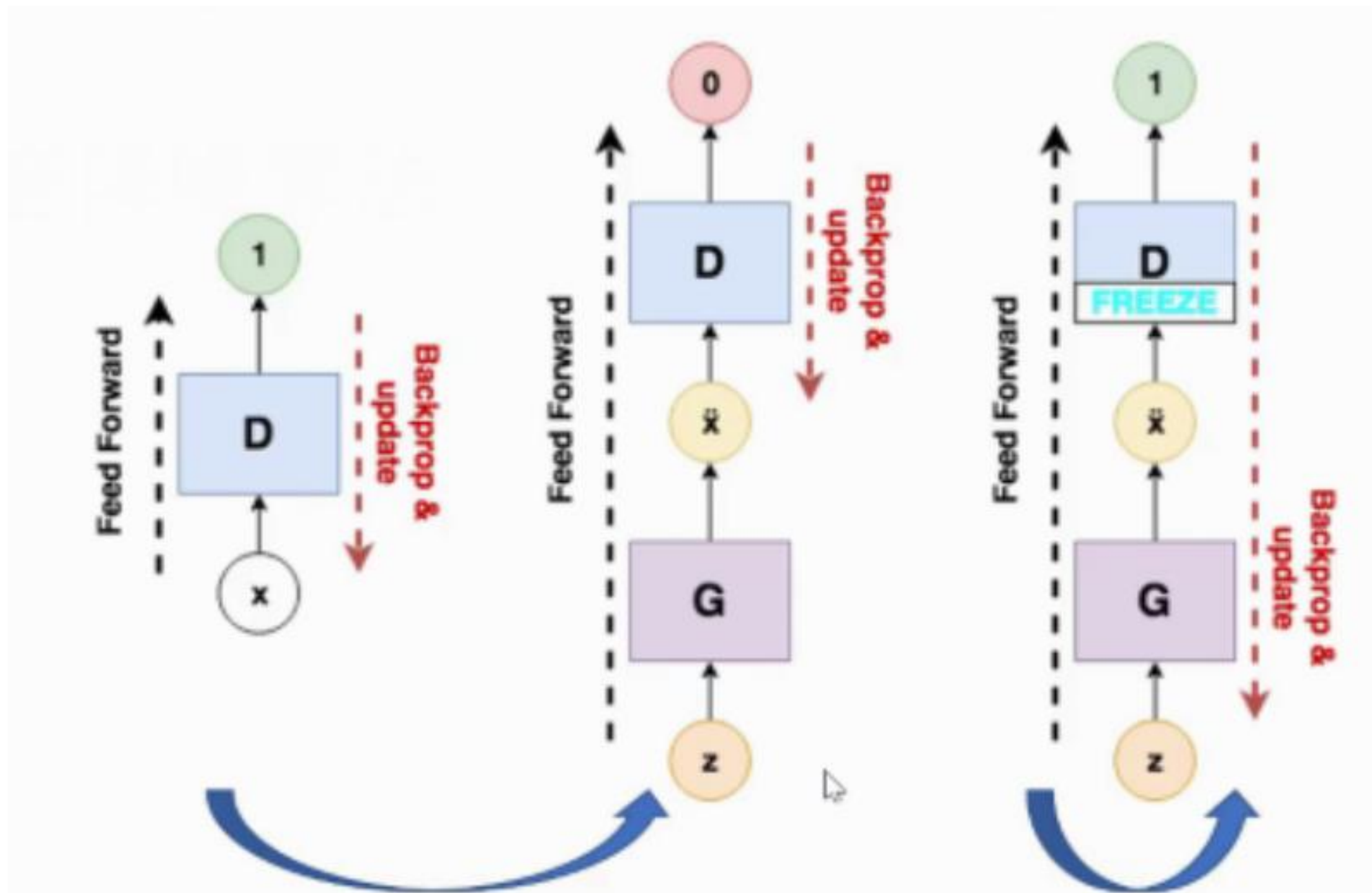
整个网络由CNN组成。

Encoder（由步长为2的1维卷积层构成）：
16384X32、8192x16、4096X32、2048X32.....
16X512、8X1024

G:Encoder-Decoder

D:结构为Encoder，接降维层，将参数降为8个

基于GAN语音降噪算法



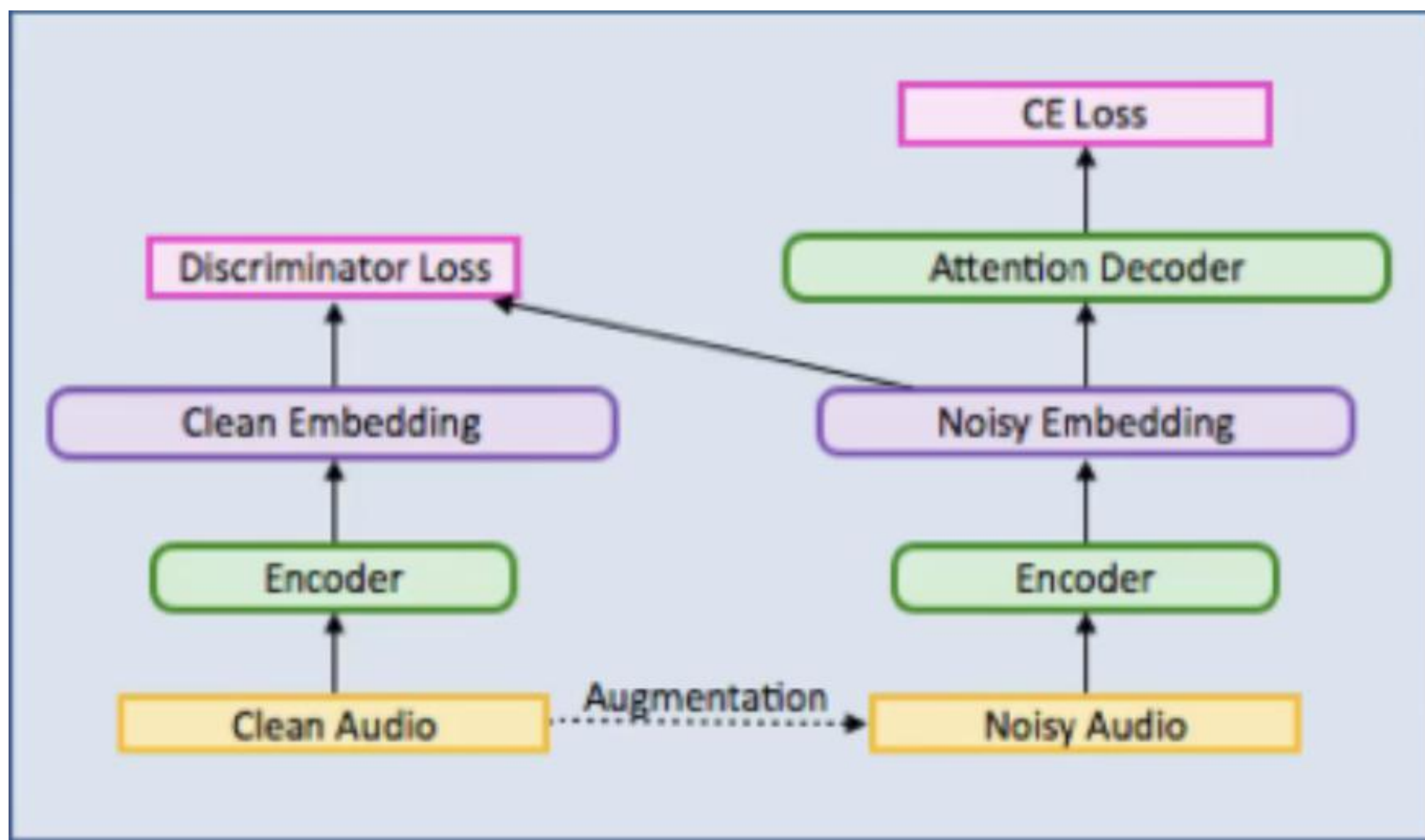
先训练一个判别器，判别器的输入分别为纯净语音信号的经过生成器增强后的信号

固定判别器，改变生成器参数号

基于GAN语音降噪算法

百度基于GAN的语音识别框架

Embedding distance:语义相似性



谢 谢！