RAWNET - 2

Rawnet-2. Rawnet modeli, uzun uca çalışan derin öğrenme tabanlı bir speaker verification modelidir. Model, ham audio veri paketleri üzerinde çalışır, bu nedenle geleneksel özellik çıkarımı ve öğrenme teknikleri kullanılarak çıkarılamayacak özellikleri öğrenebilir. →

Geliştiriciler tarafından el yardımıyla seçilen/çıkarılan özellikler, ağırlıkla bilinen saldırı türlerinin karakteristikleri göz önüne alınarak belirlendiğinden, daha önce karşılaşılmamış ve buraya özellik seçimi yapılırken hesaba katılmamış, saldırıların tespitinde yetersiz kalabilirler.

- Ham veri işleme : geleneksel özellik çıkarımı yöntemlerine bağlı kalmadan doğrudan ham veriden öğrenir.
- ResNet (residual network) tabanlı derin ağ : ResNet tabanlı evrişimli katmanlar kullanarak ham veriden hiyerarşik özellikler çıkarır.

- ResNet : residual network yani artık ağ. artık fonksiyonlar (residual function) kullanarak çok derin ağlarda ortaya çıkan bozulma (degradasyon) problemlerinin önüne geçmeye çalışır. Ağın derinliğinin artması bir noktaya kadar performansı iyileştirirken, belirli bir derinliğin üzerine çıkıldığında vanishing/exploding gradient denen bir hataya sebep olur, bu hata gradyanın 0 (sıfır) (vanishing) ya da çok büyük (exploding) olmasına sebep olur. Bu sebeple belirli bir derinlikten sonra ağın katman sayısını arttırmak eğitim ve test oranındaki hatayı arttırır. Ayrıca bu çok derinlik durumu girdi bilgisinin unutulmasına sebep olabilir.

Artık fonksiyonlar, girdi ile çıktı ile doğrudan bir ilişki kurmak yerine; çıktıyı, girdi ile çıktı arasındaki farkın fonksiyonunun bir toplamı olarak ifade eder.
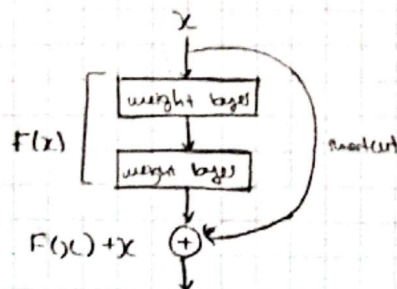
örn:
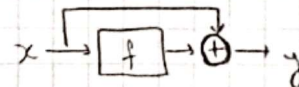
$$y = f(x) \qquad\qquad y = x + f(x)$$

Bu yöntemle girdi ağ üzerinde yayılır ve kaybolmaz. Artık ağlar, artık blokların (residual block) bir araya gelmesi ile oluşur ve bu bloklarda "skip connection" ya da "shortcut" denilen bir teknik kullanılır. Bu teknik, birbirleriyle doğrudan bağlantılı olmayan katmanlar arasında - aradaki katmanları atlayarak - kısayol bağlantıları oluşturur.

$$F(x) + x$$

- Gated Recurrent Unit (GRU) ile zaman serisi analizi: GRU katmanları, konuşmanın zaman içindeki değişimini modelleyerek konuşmacı kimliğini daha iyi yakalar.

  - GRU, RNN (recurrent neural network) sınıfına ait bir sinir ağı hücresidir. Zaman serisi verilerini işlemek için kullanılır. Uzun vadeli bağımlılıkları öğrenebilir, yani bildiklerini unutmadan yeni bilgileri hatırlayabilir. Bu özelliği sayesinde çok derin ağlarda ortaya çıkan vanishing gradient problemine de çözüm sağlar. Machine translation, speech recognition, speaker verification, text generation gibi alanlarda yaygın olarak kullanılır.

    GRU'nun iç yapısında iki kapı bulunur;
    - update gate: bilginin ne kadarının korunacağını ve aktarılacağını belirler. Kapının ürettiği değere göre geçmiş bilgiler korunur ya da yeni bilgiler ağırlık kazanır.
    - reset gate: geçmiş bilgilerin ne kadarının unutulacağını belirler.

    Özetle GRU, hücreye gelen verinin ne kadar kullanışlı olduğuna karar verir.

- SincNet, konuşma sinyallerini ham dalga formları üzerinden işleyebilen özel bir CNN (convolutional neural network) modelidir. Genellikle speaker recognition ve speech analysis alanlarında kullanılır. Diğer CNN tabanlı ses işleme modellerinden farklı olarak sabit ve öğrenebilir sinüs bazlı filtreler (sinc filters) kullanır. Bu sayede klasik CNN'e kıyasla daha az parametre alarak daha verimli çalışır.

  Geleneksel CNN'lerde ilk ~~~~~ katman için rastgele başlatılan ve zaman içinde optimize edilen kernellardan oluşan filtreler kullanılır. SincNet'de ise ilk katmandaki filtreler için sinüs bazlı fonksiyonlar (sinc) kullanılır ve bu fonksiyonların kesme (cut-off) frekansları öğrenilir.

  $$h(t, f_1, f_2) = 2f_2 \frac{\sin(2\pi f_2 t)}{2\pi f_2 t} - 2f_1 \frac{\sin(2\pi f_1 t)}{2\pi f_1 t}$$

  - f_1 ve f_2: öğrenilebilir alt ve üst kesme frekansları
  - $h(t, f_1, f_2)$: bant geçiren (band-pass) bir filtre üretir.