

DL FINAL PROJESİ

Conformer mimarisini kullanarak MIDI müzik üretimi



Tabib Hasanzada
15.12.2025



Problem Tanımı

1

Transformer & CNN

- CNN mimarileri, konvolüsyon yapısı sayesinde yerel motifleri ve anlık dokusal detayları (Local Context) yakalamada oldukça başarılıdır
- Transformer modelleri, Self-Attention mekanizması ile uzun vadeli bağımlılıkları ve parçanın genel yapısını (Global Context) öğrenmede üstündür

2

Conformer

- Conformer, bu iki mimarinin en güçlü yönlerini yapısı içinde birleştirerek hibrit bir çözüm sunar
- Bu yapı, CNN modülü ile notalar arası geçişleri (yerel) işlerken, Attention modülü ile eserin bütünlüğünü (global) koruyarak her ikisini öğrenmeyi hedefler.

3

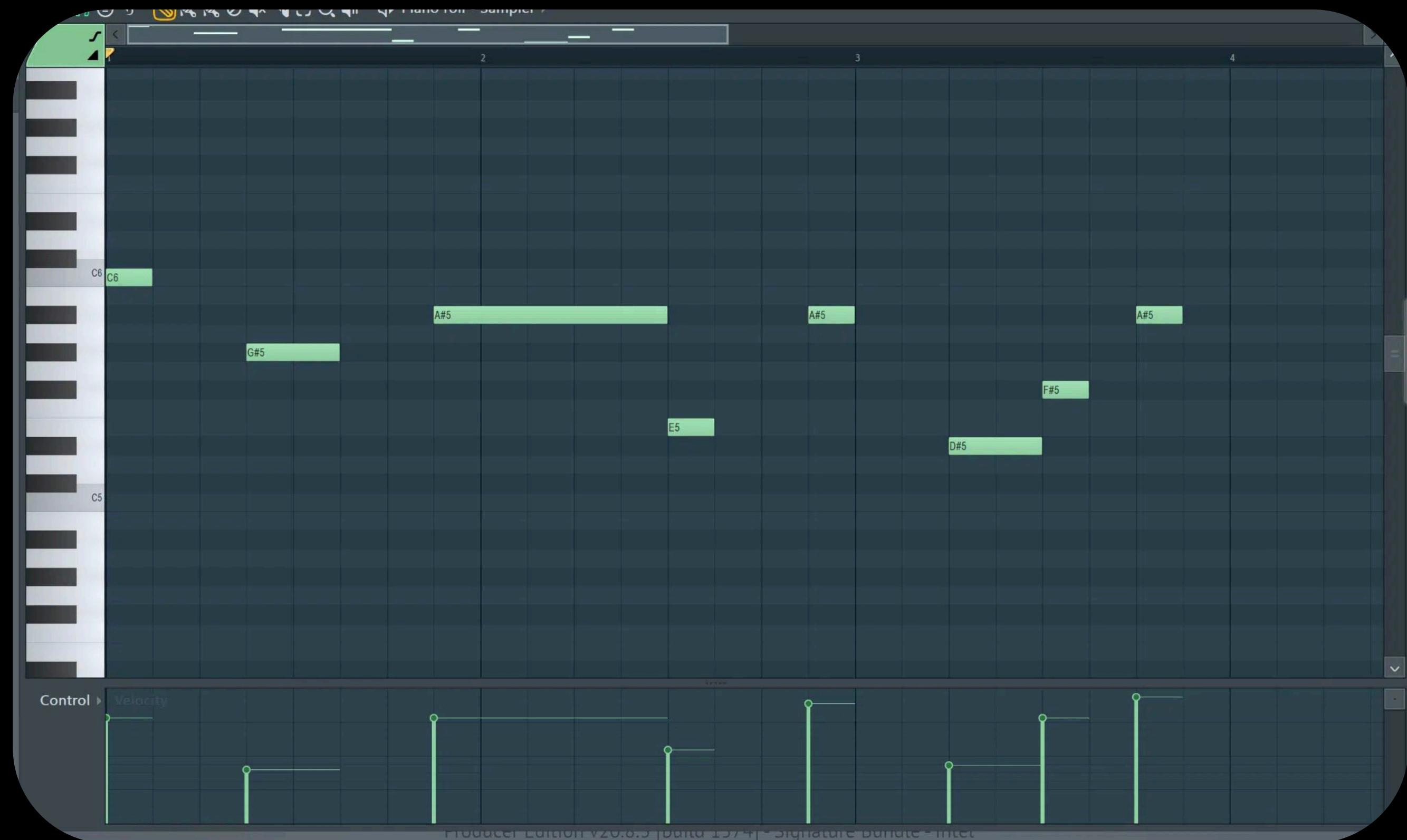
Music Transformer

- Güncel SOTA modelleri (örn. Google Magenta) music transformer kullanmaktadır. Music Transformer modelleri, standart attention yerine Relative Self-Attention (Göreceli Dikkat) kullanmaktadır. Bu mekanizma, notaların mutlak sırasına değil, birbirlerine olan mesafesine odaklanarak bir motifin şarkının herhangi bir yerinde tanınmasını (Translation Invariance) sağlar

4

Conformer vs SOTA

- Bu projede Conformer mimarisinin hem yerel hem de global bağlamı aynı anda işleme yeteneğinin, sembolik müzik (MIDI) üretimindeki başarısını ve tutarlığını araştırmak hedeflenmiştir.



MİDİ?

- MIDI (Musical Instrument Digital Interface), sesi dalga formu (audio/ses kaydı) olarak değil, dijital talimatlar seti olarak temsil eden bir protokoldür.
- İçerik: Bir ses dosyası yerine; "Hangi nota?", "Ne zaman?", "Ne kadar süreyle?" ve "Hangi şiddetle (Velocity)? basılacağı bilgisini taşır.
- Veri Temsili: Sembolik bir veri yapısı olduğu için, tipki kelimeler gibi tokenleştirilerek (tokenization) dil modellerinde (NLP) işlenmeye son derece uygundur.

Yarım Feed-Forward (FFN): Girdi önce normalize edilir, ardından FFN katmanından geçer. Çıktı 0.5 ile çarpılıp (yarım adım) girdiye eklenir (Residual Connection).

Self-Attention Modülü: Global (uzun vadeli) ilişkileri yakalar. Multi-Head Attention kullanılır ve Causal Mask uygulanır (gelecek notaları görmemesi için).

Konvolusyon: Burası modelin "ritim ve yerel desen" yakaladığı yerdir.

Pointwise Conv: Boyutu genişletir.

GLU (Gated Linear Unit): Kapılama mekanizması ile önemli bilgiyi süzer.

Depthwise Conv (1D): Zaman ekseninde (notalar üzerinde) kayar pencere ile işlem yapar.

Pointwise Conv: Boyutu tekrar eski haline getirir.

Yarım Feed-Forward (FFN): Bloğun sonunda tekrar bir FFN katmanı vardır. Çıktı yine 0.5 ile çarpılıp eklenir.

Layer Normalization: Son çıkış normalize edilir.



Model Mimarisi

Projede kullanılan Conformer bloğu, Google'in orijinal makalesindeki yapının Müzik Üretimine (Causal Generation) uyarlanmış halidir. Standart Transformer'dan en büyük farkı Macaron (Sandviç) yapısıdır.

embedding_size: 512
conformer_layers: 8
attention_head: 8
sequence_length: 1024

Veri Setleri

Bach Chorales

J.S. Bach'a ait 4 sesli eserler. Kısa sekanslar, katı armonik kurallar ve homofonik doku.

DATA

382

TOKENS

1,220,000

POP909

909 adet popüler şarkının profesyonel piyano düzenlemesi

DATA

909

TOKENS

900,000

MAESTRO

Uluslararası Piyano Yarışması'ndan (Piano-e-Competition) alınan virtüöz performansları.

DATA

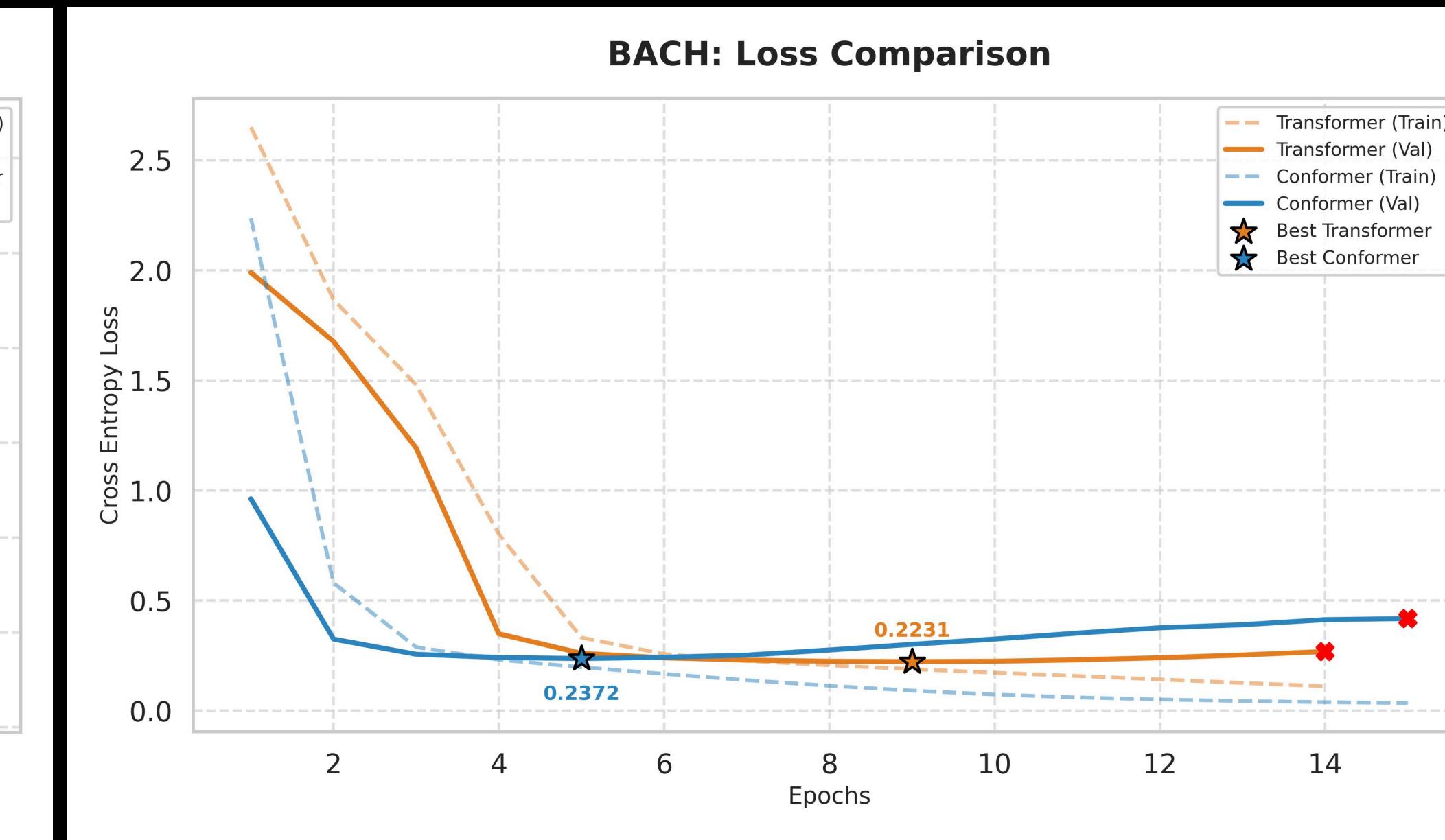
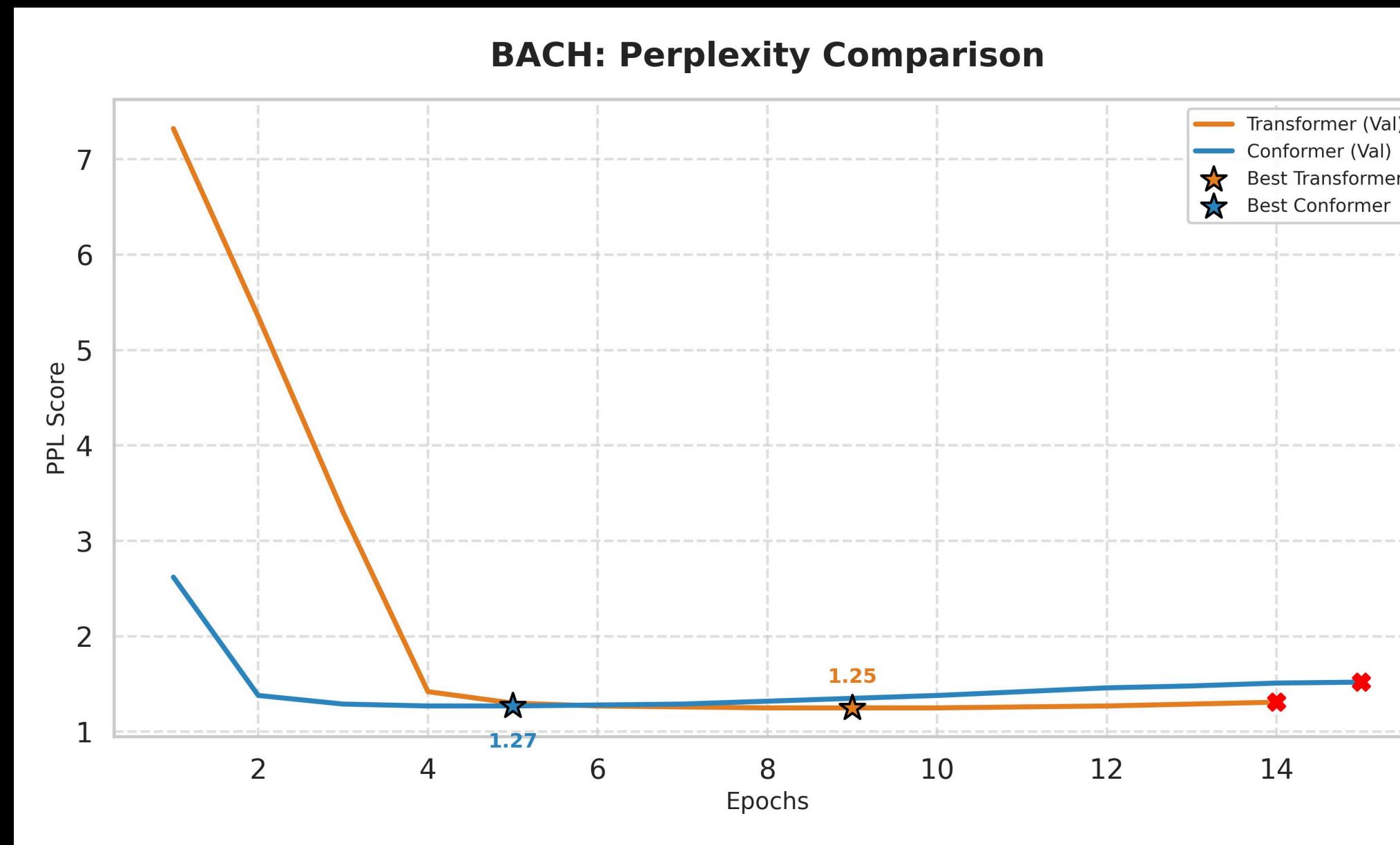
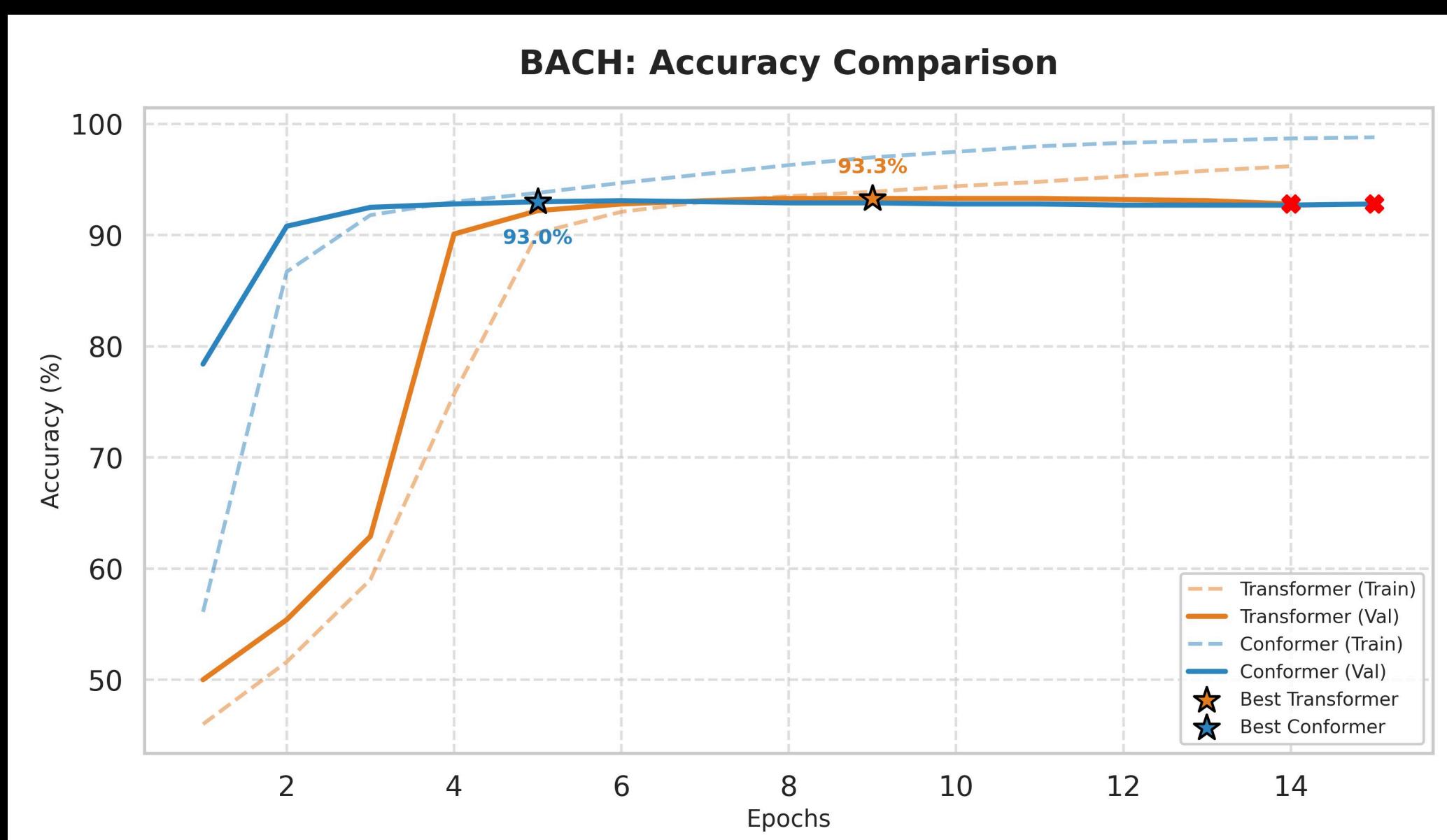
1276

TOKENS

25,600,000

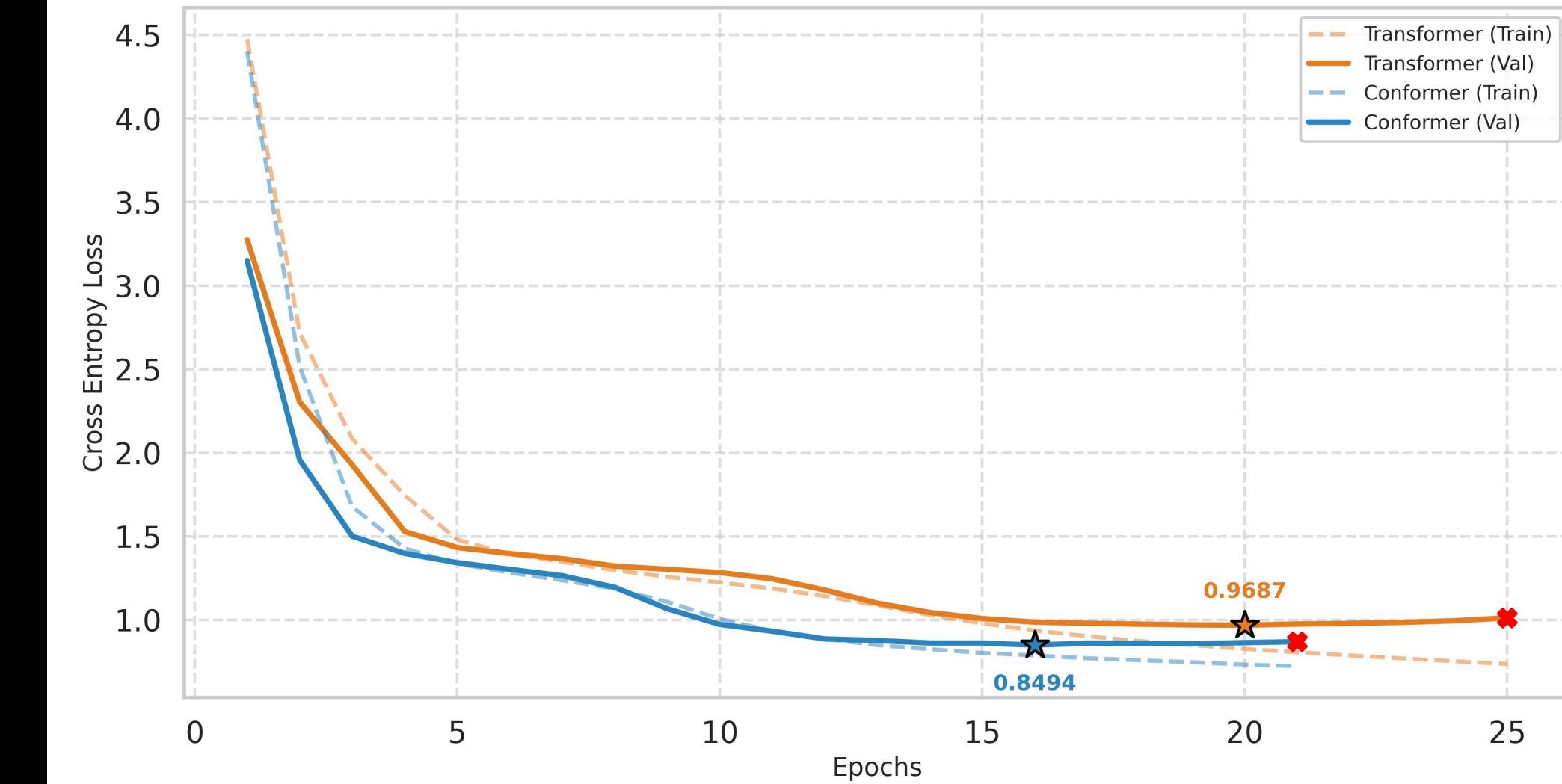


Bach

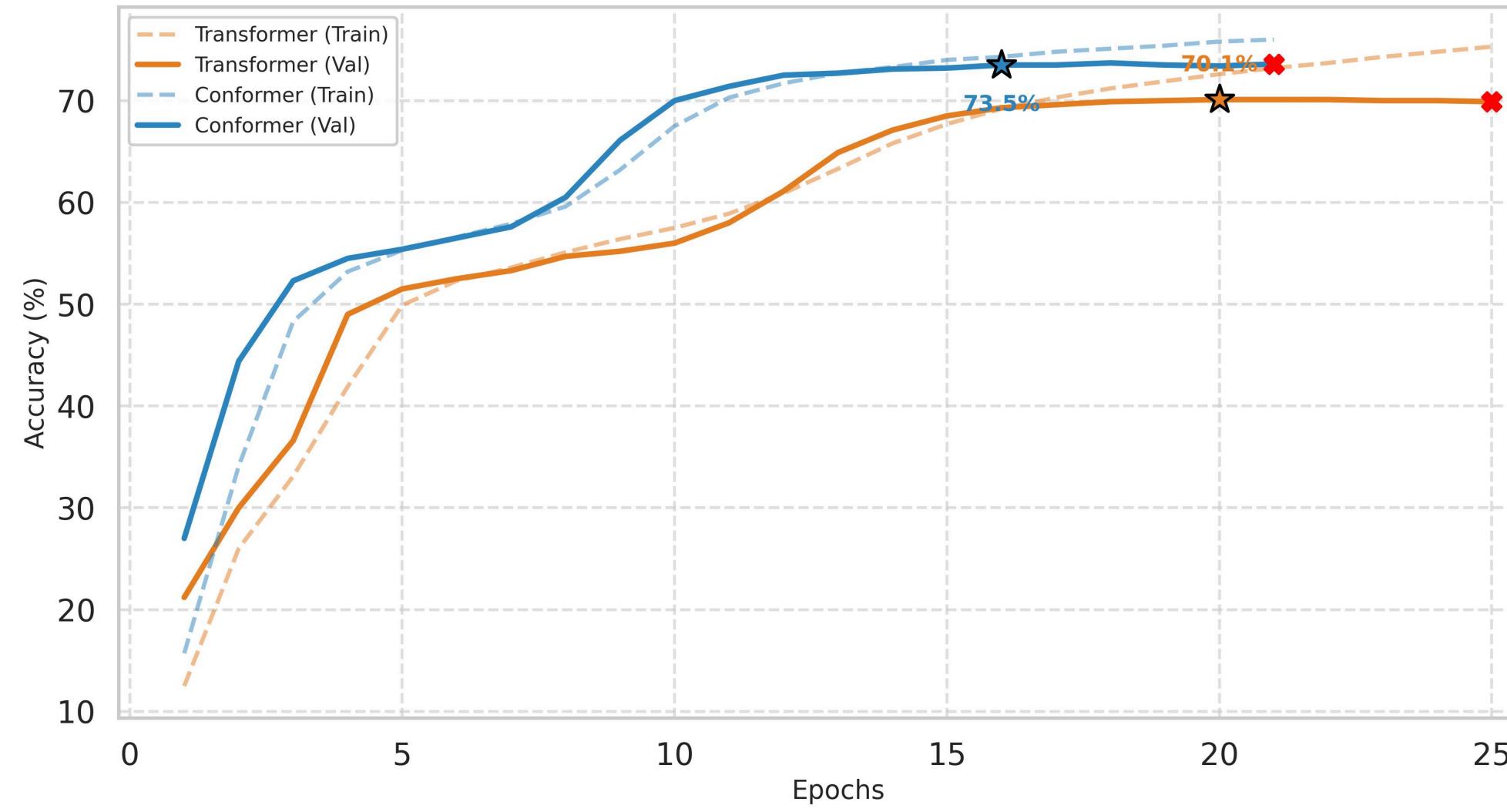


POP909

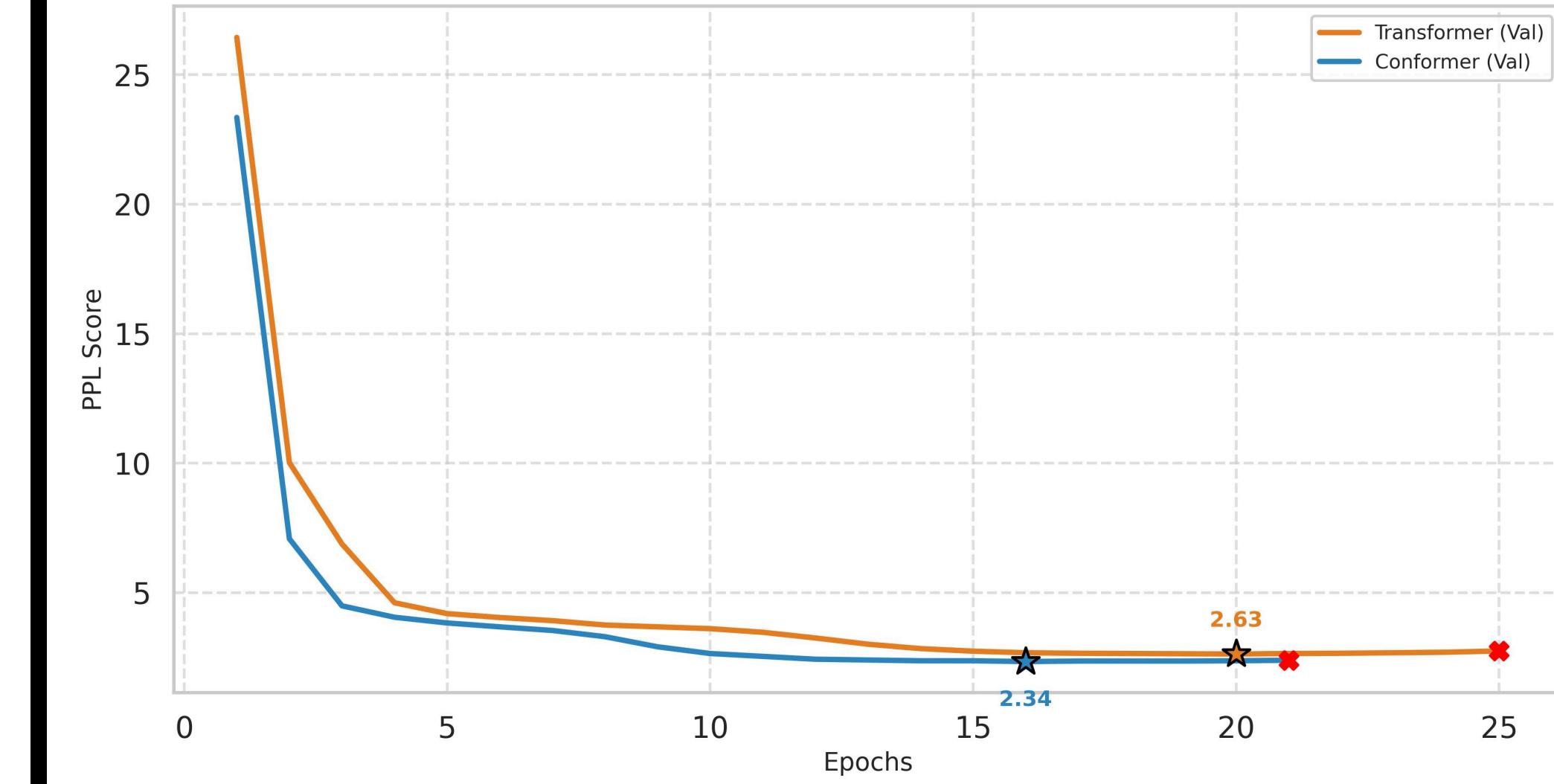
POP909: Loss Comparison



POP909: Accuracy Comparison



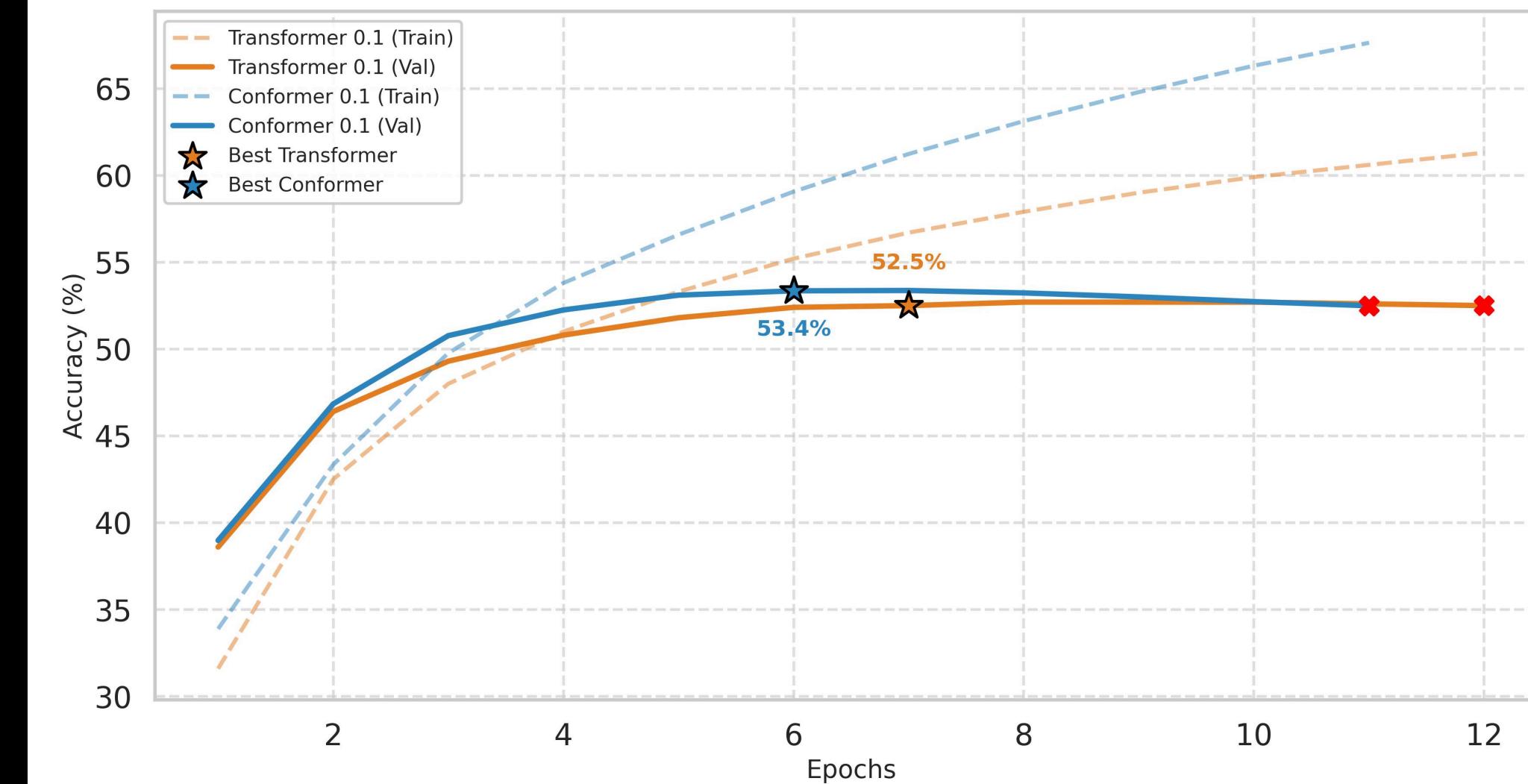
POP909: Perplexity Comparison



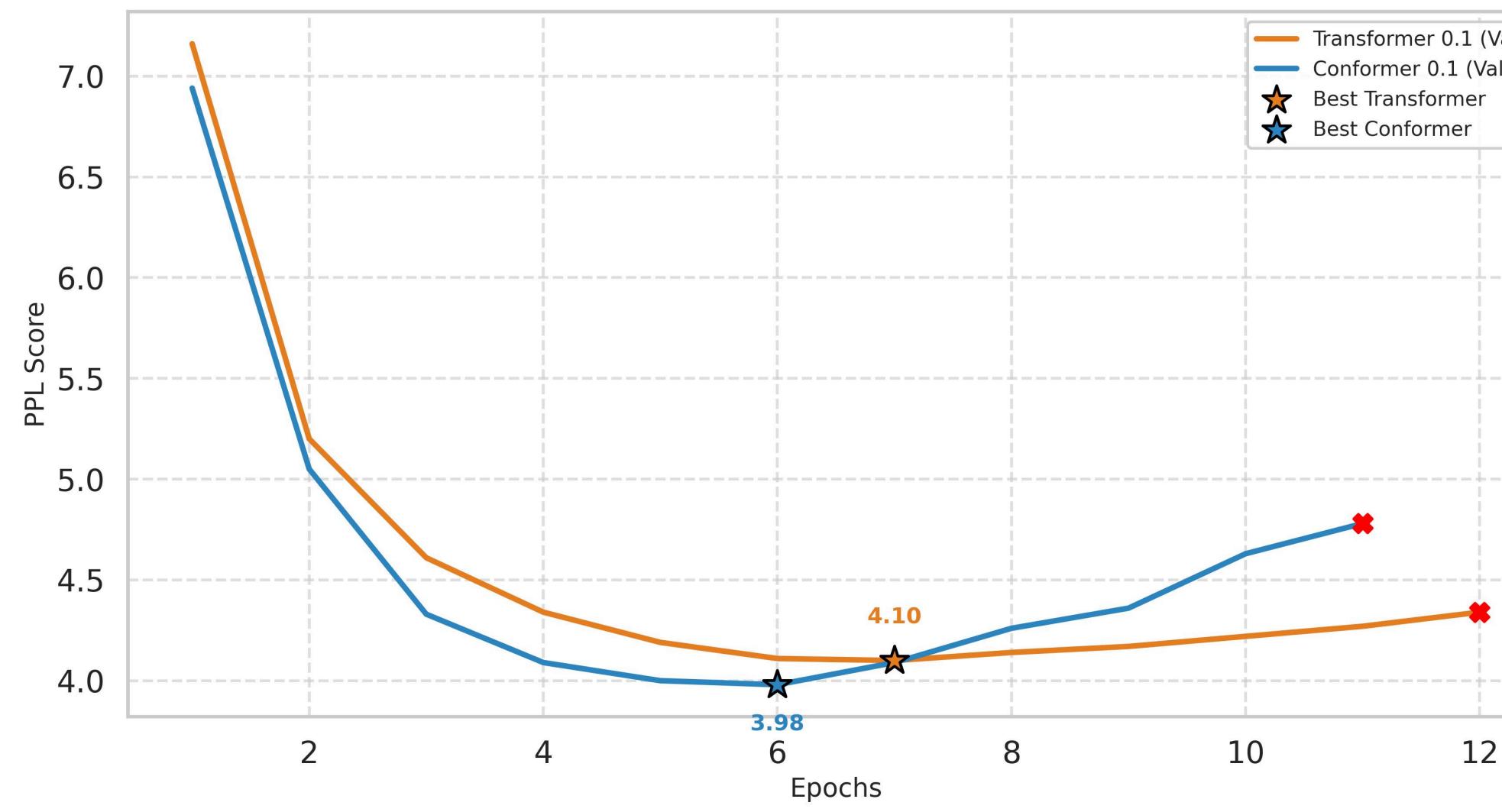
MAESTRO

0.1 dropout

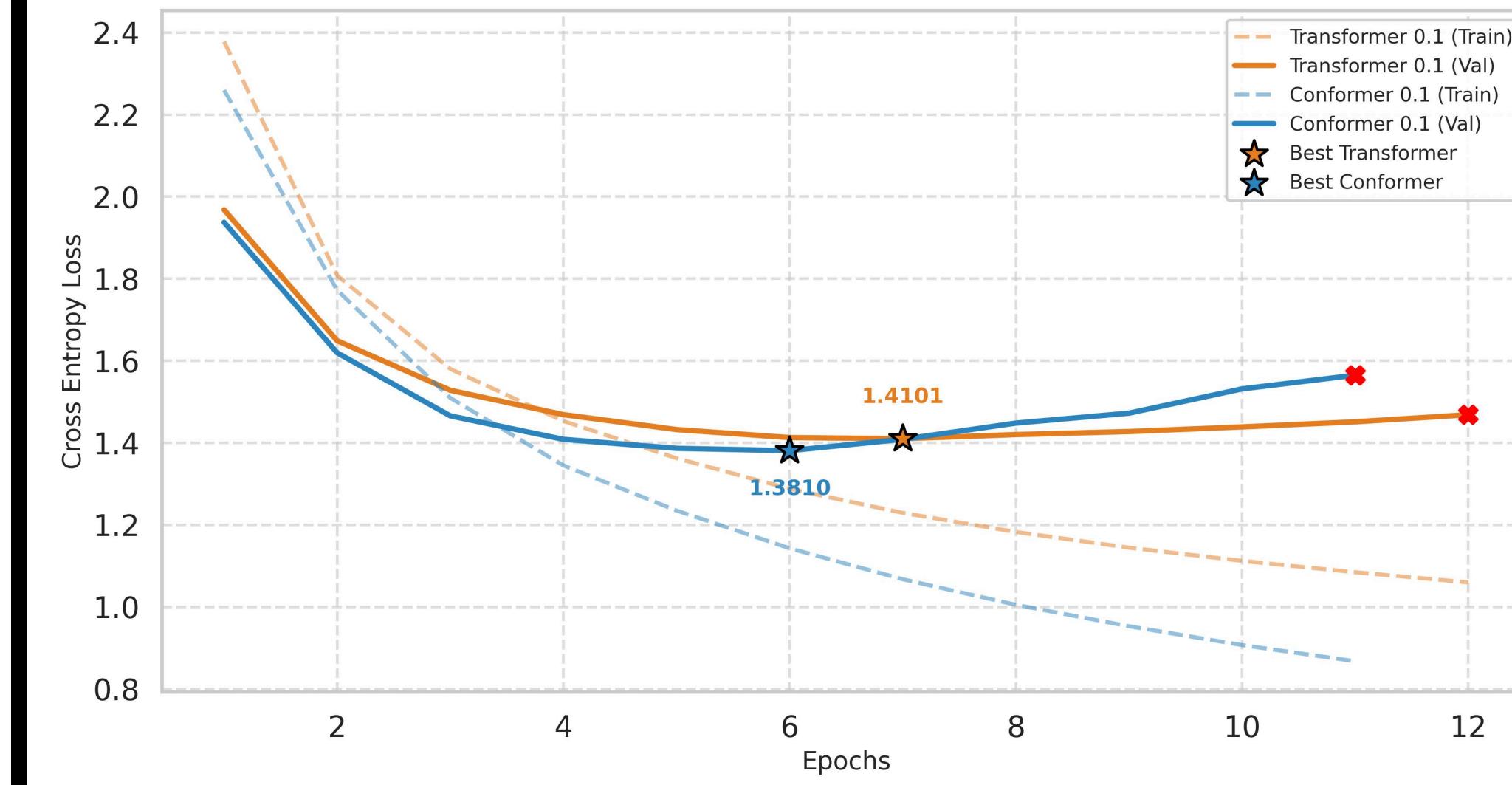
Comparison (0.1 Dropout): Accuracy



Comparison (0.1 Dropout): Perplexity



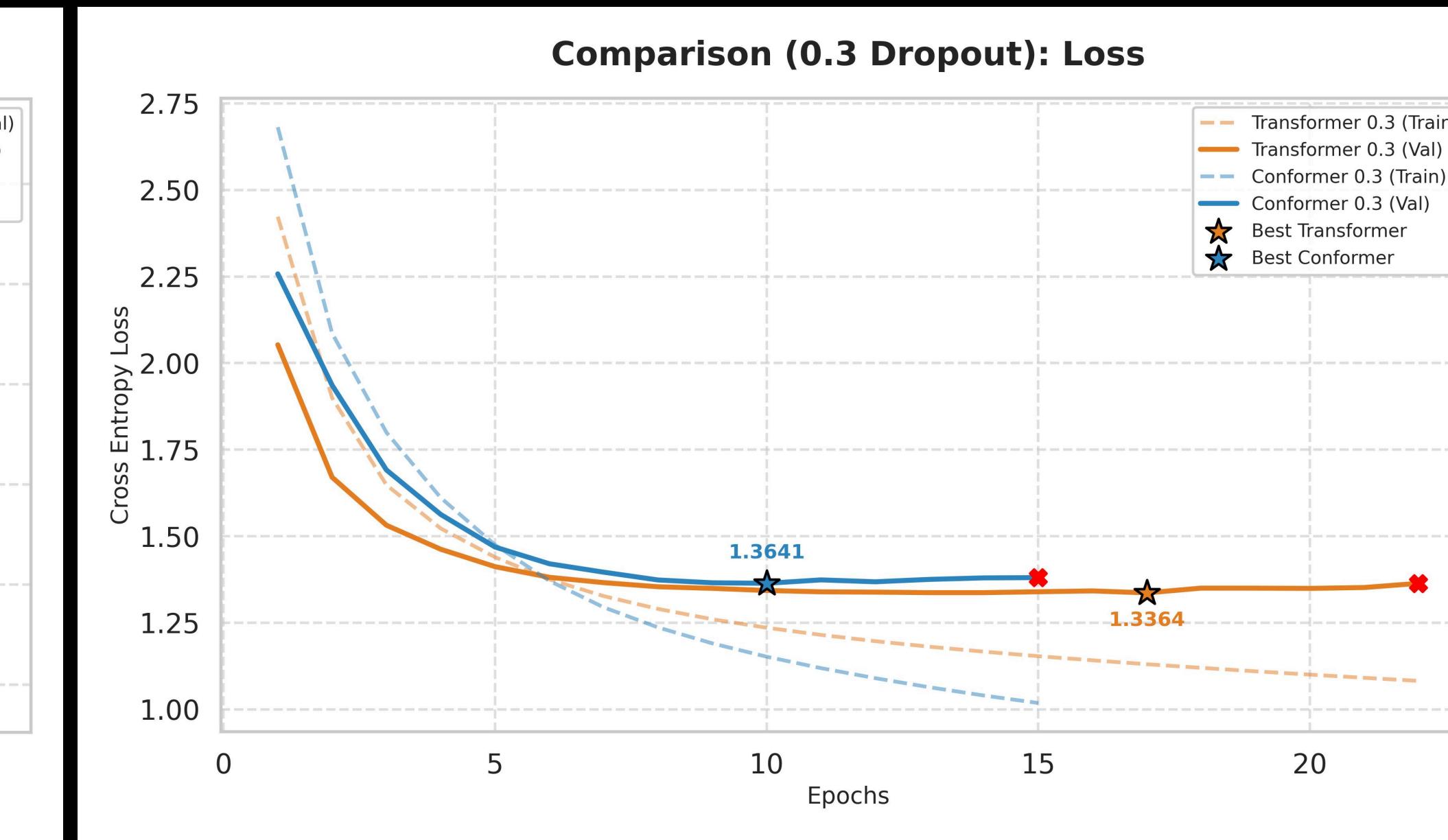
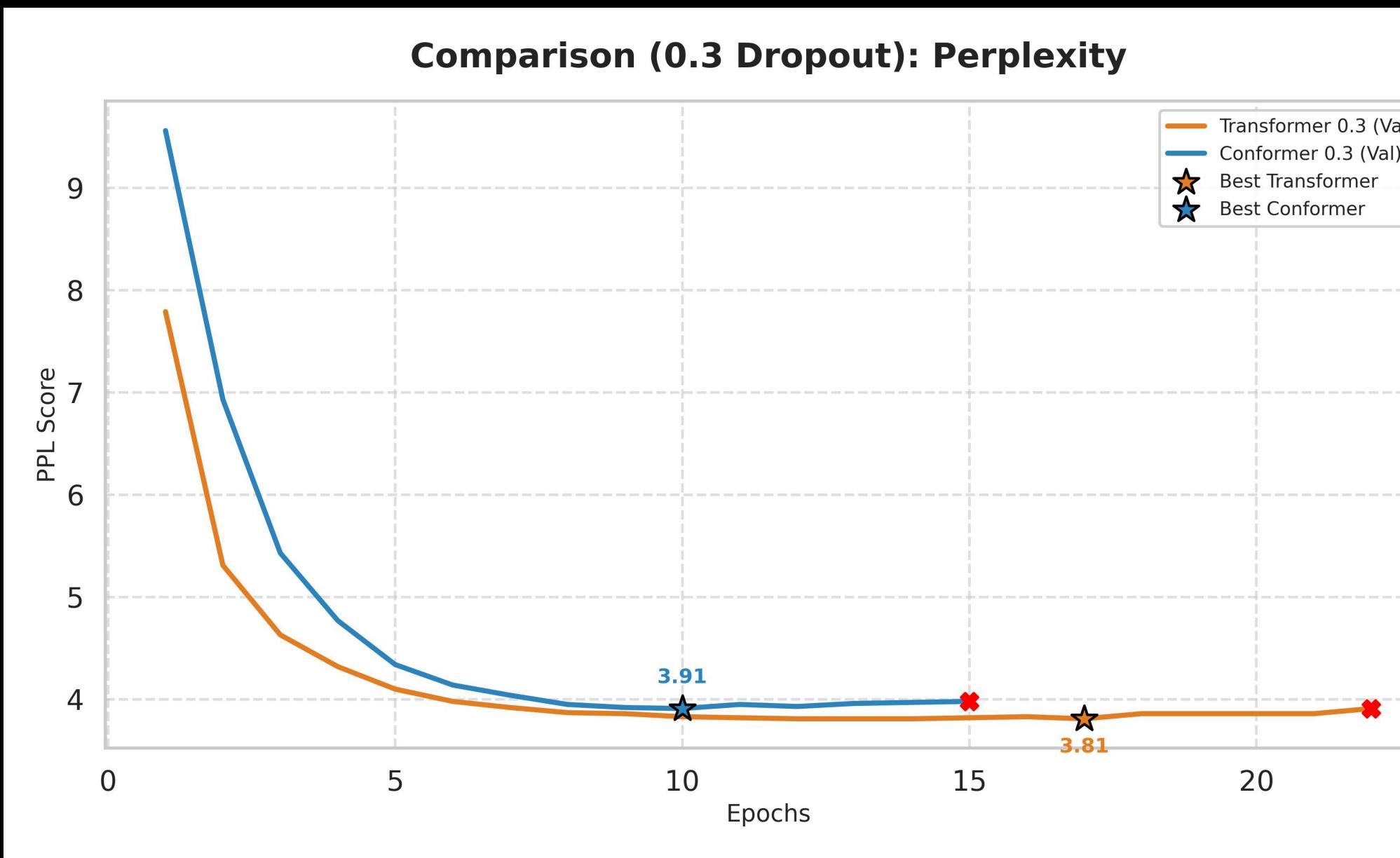
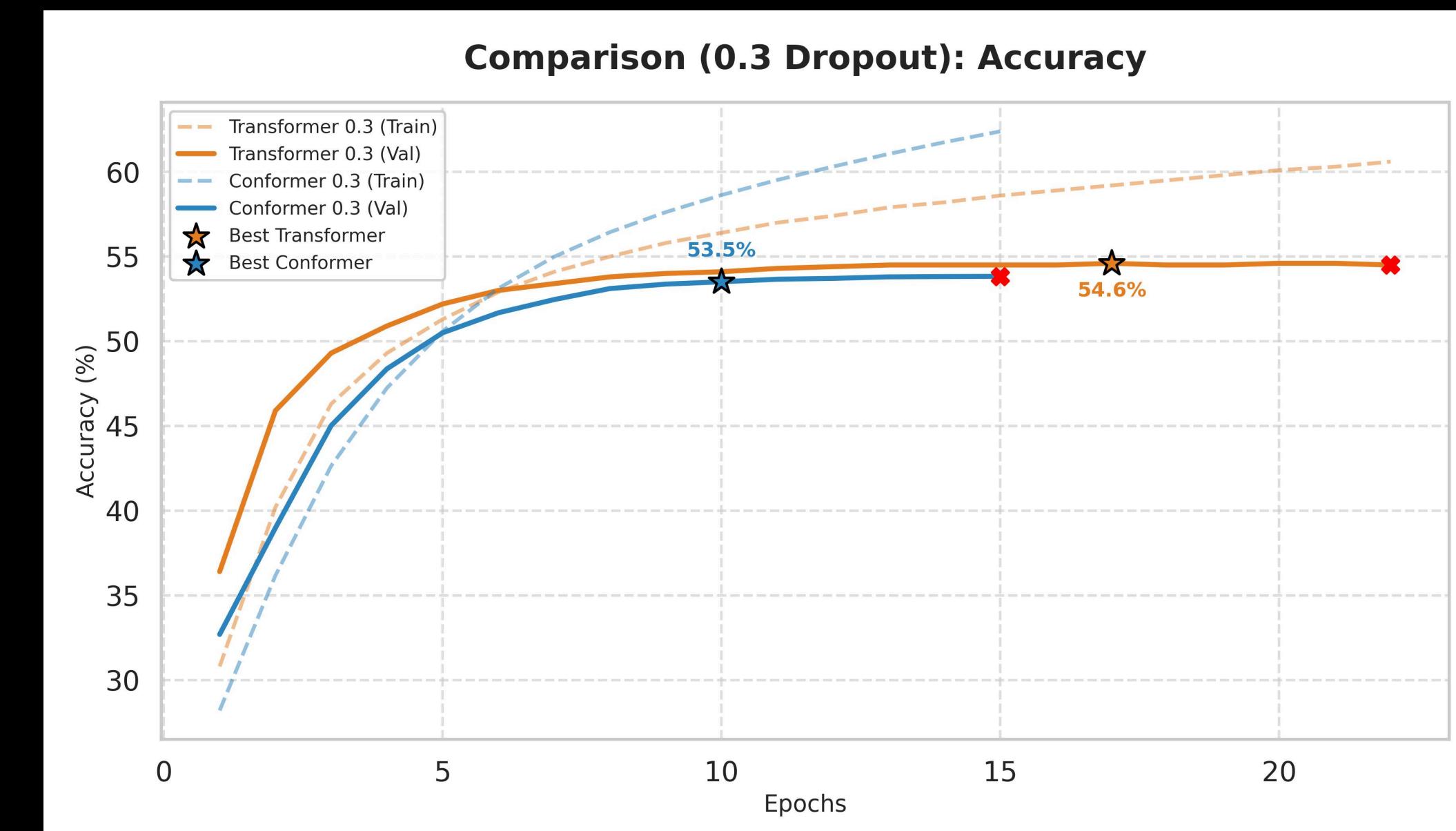
Comparison (0.1 Dropout): Loss



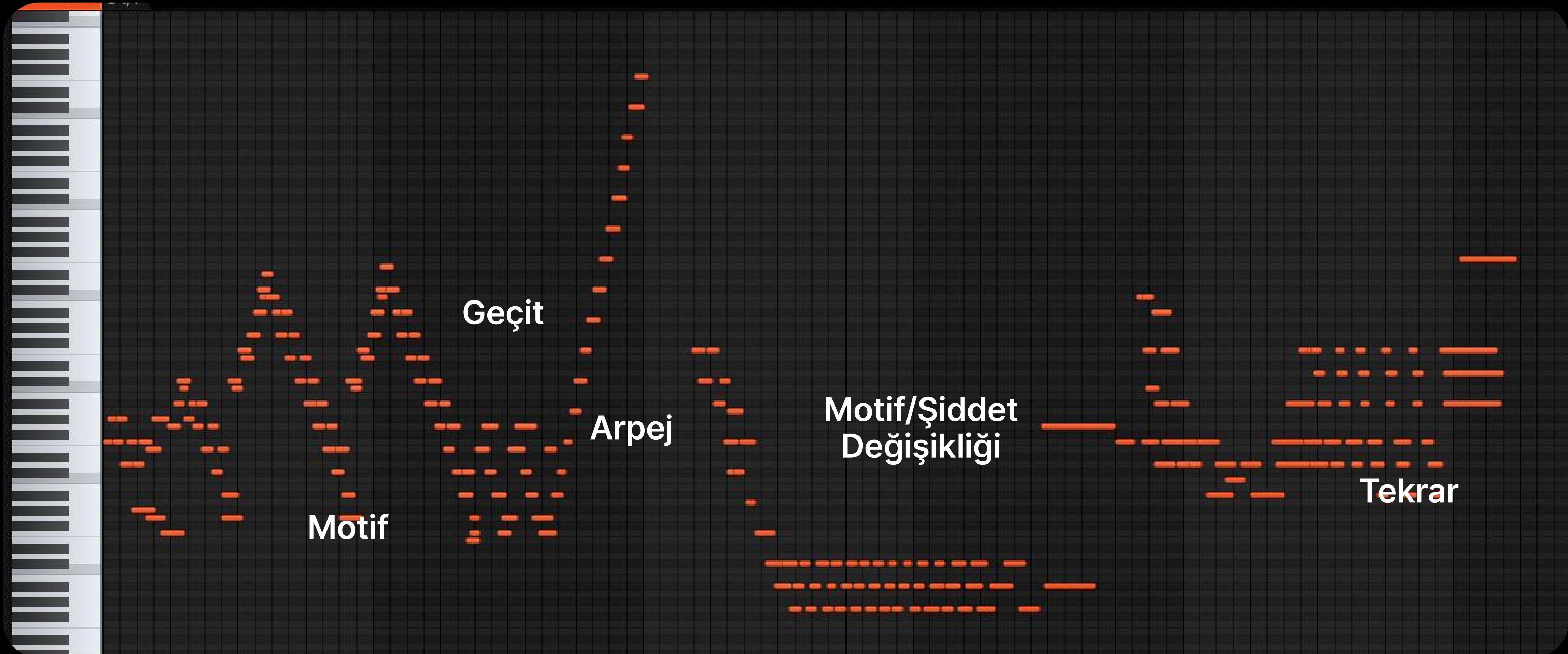
MAESTRO

0.3 dropout

Modelin ezberlemesini önlemek ve öğrenme kapasitesini korumak arasındaki dengeyi bulmak için farklı dropout değerleri (0.1 ve 0.3) denenmiştir.



Conformer mimarisiyle oluşturulmuş örnek midi ve onun kullandığı teknik detaylar



SONUÇLAR

Üstünlük ve Eksiklikler

Yapı

Model özellikle MAESTRO veri setinde zamanla farklı kararlar vererek oluşturulan müziğin yapısını daha da geliştirdi

Piano Teknikleri

Model MAESTRO veri setindeki çıktılarından da gözüktüğü üzere piano tekniklerini öğrendi

Ritm

Model hem gereği zaman 4/4-lük ritmi korudu, hem MAESTRO veri setinde farklı ritmler ve seslenmeler denedi

Basım gücü

Model notaların ne zaman yavaş ne zaman daha güçlü basımını da öğrendi

Melodi

Model çoğu zaman belirgin melodiyi üretmekte zorluk çekti

Nota kaymaları

Model çoğu zaman doğru notalarda, gamada müzik üretse de bazen artifactlar oluşturma

Tekrar

Model bazen döngülere girebiliyor

SONUÇLAR

Tartışma ve Geliştirme

Model yerel contexti daha iyi anladı

Global context-i daha iyi öğrenmesi için modelin self-attention kısmı daha geliştirilebilir

Daha dengeli veri setiyle eğitilebilir

Daha büyük ve dengeli bir müzik türüne odaklanan orta zorlukta modelle daha iyi eğitilebilir

Daha uzun sequence length ve context window

MAESTRO veri setinin tümünü daha uzun seq_length ve context windowla eğitilip daha iyi sonuçlar elde edilebilir

Dinlediğiniz için
teşekkürler

