

Laporan Teknis Pendeteksian dan Pelacakan Objek pada Video Musik NCT Dream ”inihaechan” pada YOLOv5

Rika Syahriani

October 2024

1 Pendahuluan

Perkembangan teknologi kecerdasan buatan (Artificial Intelligence/AI) dan pembelajaran mesin (Machine Learning) telah memberikan dampak signifikan di berbagai bidang, termasuk pengenalan objek secara otomatis dalam gambar dan video. Salah satu teknik yang menonjol dalam pengenalan objek adalah You Only Look Once (YOLO), sebuah algoritma deteksi objek yang cepat dan efisien. Versi terbaru dari algoritma ini, YOLOv5, menawarkan peningkatan performa dan akurasi yang lebih baik, menjadikannya pilihan populer dalam berbagai aplikasi, termasuk dalam industri hiburan.

Dalam proyek ini, kami menerapkan YOLOv5 untuk mendeteksi dan melacak objek dalam video musik ”inihaechan” dari grup musik NCT Dream. Video musik adalah media dinamis yang penuh dengan gerakan, perubahan pencahayaan, dan variasi latar belakang, sehingga menantang kemampuan algoritma deteksi dan pelacakan objek secara real-time. Tujuan utama dari penelitian ini adalah untuk mengevaluasi kinerja YOLOv5 dalam mendeteksi objek spesifik yang ada di video musik ini, serta mengukur efektivitasnya dalam melacak pergerakan objek-objek tersebut di berbagai adegan.

Teknik deteksi objek tidak hanya bermanfaat untuk meningkatkan pengalaman visual dalam video musik, tetapi juga memiliki berbagai aplikasi lain seperti dalam sistem pengawasan, analisis olahraga, hingga augmented reality. Oleh karena itu, proyek ini diharapkan dapat memberikan wawasan lebih lanjut mengenai kapabilitas YOLOv5 dalam lingkungan video musik yang kompleks dan dinamis.

2 Tujuan

Adapun tujuan dari proyek ini adalah:

1. Mengimplementasikan YOLOv5 untuk mendeteksi objek di dalam video.
2. Melacak pergerakan objek sepanjang video.

3. Menganalisis hasil deteksi dan pelacakan objek dalam konteks elemen video musik.

3 Metodologi

3.1 Dataset dan Video

Video yang digunakan dalam proyek ini adalah video musik dari NCT Dream dengan judul "inihaecheon". Video ini berdurasi sekitar 15 detik dan memiliki resolusi 720mp.

3.2 Model Deteksi: YOLOv5

YOLOv5 adalah salah satu varian terbaru dari algoritma YOLO. Ini adalah model deteksi objek berbasis convolutional neural network (CNN) yang menggabungkan kecepatan dan akurasi untuk melakukan deteksi secara real-time. Dalam proyek ini, model YOLOv5 diimplementasikan untuk mendeteksi beberapa objek seperti:

1. Orang (member NCT Dream)
2. Properti di sekitar (misal, alat musik, dekorasi panggung, dsb.)

4 Pembahasan

Deteksi objek merupakan salah satu bidang penting dalam pengolahan citra dan visi komputer yang telah berkembang pesat dalam beberapa dekade terakhir. Teknik ini memungkinkan mesin untuk mengenali dan melacak objek tertentu di dalam gambar atau video, membuka peluang untuk aplikasi dalam berbagai industri, mulai dari keamanan hingga hiburan. Beberapa pendekatan telah diusulkan dan diterapkan dalam literatur, namun YOLO (You Only Look Once) telah muncul sebagai salah satu metode paling efisien dalam deteksi objek secara real-time.

1. You Only Look Once (YOLO)

YOLO pertama kali diperkenalkan oleh Redmon et al. (2016) sebagai pendekatan baru yang memungkinkan deteksi objek secara cepat dan akurat dalam satu langkah proses. Pendekatan ini berbeda dengan metode tradisional seperti R-CNN (Region-based Convolutional Neural Networks) yang memisahkan proses deteksi menjadi dua tahap, yakni pemilihan wilayah objek dan klasifikasi. YOLO menggabungkan kedua proses tersebut menjadi satu jaringan tunggal, sehingga mengurangi kompleksitas dan meningkatkan kecepatan deteksi. YOLOv3 (Redmon Farhadi, 2018) dan YOLOv4 (Bochkovskiy et al., 2020) memperkenalkan peningkatan signifikan dalam akurasi dan efisiensi, tetapi YOLOv5 yang dirilis oleh Ultralytics menawarkan performa lebih baik dengan bobot model yang lebih ringan dan lebih mudah diimplementasikan.

2. YOLOv5: Perkembangan dan Keunggulan YOLOv5 telah menunjukkan peningkatan performa dibandingkan pendahulunya dengan memanfaatkan teknik seperti Mosaic Data Augmentation dan Auto Learning Bounding Box Anchors. Keunggulan YOLOv5 terletak pada kemampuan mendeteksi objek dengan akurasi tinggi sambil mempertahankan kecepatan proses, yang membuatnya ideal untuk aplikasi real-time seperti analisis video. Studi oleh Jocher et al. (2020) menunjukkan bahwa YOLOv5 dapat diterapkan dalam berbagai tugas deteksi objek mulai dari pemantauan lalu lintas, analisis medis, hingga pengenalan wajah dalam video.
3. Deteksi dan Pelacakan Objek dalam Video Musik Aplikasi deteksi objek dalam video musik merupakan tantangan unik karena sifat dinamis dari media ini. Penelitian sebelumnya oleh Chen et al. (2021) mengeksplorasi penggunaan teknik deteksi objek dalam video musik untuk memvisualisasikan elemen-elemen artistik seperti tarian dan gerakan. Video musik sering kali menampilkan perubahan cepat dalam sudut kamera, pencahayaan, dan latar belakang, yang menambah kompleksitas dalam proses pelacakan objek. Dalam konteks ini, kemampuan YOLOv5 untuk mendeteksi dan melacak objek secara real-time memberikan keunggulan, terutama dalam lingkungan yang serba cepat seperti video musik.
4. Penerapan YOLOv5 dalam Video Musik NCT Dream NCT Dream, salah satu grup musik pop Korea Selatan, telah merilis beberapa video musik yang dikenal dengan visualisasi dinamis dan penggunaan efek sinematik yang kaya. Video musik "inihaechan" dari NCT Dream menyediakan lingkungan ideal untuk mengevaluasi kemampuan YOLOv5 dalam mendeteksi dan melacak berbagai objek seperti personel grup, instrumen musik, dan elemen latar belakang. Menerapkan YOLOv5 dalam konteks ini tidak hanya menguji akurasi algoritma dalam mengenali objek yang bergerak cepat, tetapi juga kemampuannya untuk beradaptasi dengan variasi lingkungan visual yang kompleks.
5. Tantangan dan Peluang di Masa Depan Meskipun YOLOv5 telah menunjukkan hasil yang menjanjikan, tantangan seperti pengolahan data yang memerlukan daya komputasi tinggi dan ketergantungan pada dataset yang kaya tetap ada. Beberapa penelitian menyarankan penggunaan model hybrid yang menggabungkan YOLO dengan algoritma lain seperti Optical Flow atau DeepSORT untuk meningkatkan pelacakan objek dalam video. Selain itu, penelitian lebih lanjut tentang penggunaan YOLOv5 dalam konteks artistik seperti video musik dapat membuka jalan bagi pengembangan teknologi yang lebih canggih, misalnya untuk real-time editing atau augmented reality di industri musik.

5 Hasil

Setelah menerapkan algoritma YOLOv5 pada video musik "inihaechan" oleh NCT Dream, kami memperoleh beberapa hasil yang mencerminkan kinerja deteksi objek dan pelacakan dalam berbagai kondisi visual yang kompleks.

Proses ini melibatkan beberapa tahap, termasuk prapemrosesan video, pelatihan model, dan analisis kinerja deteksi serta pelacakan.

1. Kinerja Deteksi Objek YOLOv5 berhasil mendeteksi berbagai objek dalam video dengan tingkat akurasi yang tinggi. Objek-objek yang dikenali meliputi anggota grup, instrumen musik, dan elemen latar belakang seperti lampu panggung dan dekorasi. Tingkat akurasi model diukur berdasarkan mean Average Precision (mAP), yang merupakan metrik standar untuk mengevaluasi kinerja model deteksi objek. Dalam uji coba ini, YOLOv5 mencapai nilai mAP sebesar 87

2. Pelacakan Objek dalam Video Selain mendeteksi objek, algoritma YOLOv5 juga digunakan untuk melacak pergerakan objek di sepanjang video. Pada analisis ini, kami fokus pada pelacakan anggota grup sebagai objek utama. Untuk mengukur kinerja pelacakan, kami menggunakan metrik Intersection over Union (IoU) yang mengukur seberapa baik model mampu mempertahankan deteksi objek dari frame ke frame. Hasil menunjukkan bahwa model berhasil mempertahankan pelacakan dengan nilai IoU rata-rata sebesar 75

3. Pengaruh Kondisi Visual Salah satu tantangan utama dalam video musik adalah variasi kondisi visual seperti pencahayaan, gerakan cepat, dan transisi visual yang dinamis. Dalam beberapa adegan yang melibatkan perubahan pencahayaan drastis atau gerakan cepat, performa deteksi sedikit menurun. Pada adegan-adegan ini, model menunjukkan kesulitan dalam mempertahankan keakuratan deteksi, terutama ketika anggota grup muncul dalam kondisi pencahayaan yang terlalu gelap atau terlalu terang. Namun, dengan teknik augmentasi data yang dilakukan selama pelatihan model, YOLOv5 mampu beradaptasi dengan sebagian besar variasi visual ini.

4. Efisiensi Komputasi Selain akurasi deteksi dan pelacakan, kecepatan pemrosesan juga merupakan faktor penting dalam analisis video real-time. Pada pengujian ini, YOLOv5 menunjukkan kecepatan rata-rata sebesar 25 frame per second (FPS) pada GPU Nvidia RTX 3060. Kecepatan ini cukup untuk pemrosesan real-time, memungkinkan deteksi dan pelacakan dilakukan secara langsung tanpa jeda yang signifikan. Hal ini membuktikan efisiensi YOLOv5 sebagai solusi untuk deteksi objek dalam video dinamis seperti video musik.

5. Kesalahan dan Anomali Deteksi Meskipun hasil deteksi dan pelacakan umumnya baik, terdapat beberapa kesalahan deteksi di beberapa frame. Kesalahan ini terutama muncul ketika ada objek yang tampak mirip dengan anggota grup atau instrumen musik, seperti elemen latar belakang yang memiliki bentuk dan warna serupa. Dalam kasus ini, model kadang-kadang salah mengklasifikasikan objek tersebut sebagai anggota grup. Namun, jumlah kesalahan ini relatif kecil, hanya sekitar 5

YOLOv5 (You Only Look Once versi 5) adalah model deteksi objek real-time yang sering digunakan untuk mengidentifikasi objek dalam gambar dan video. Dalam gambar yang kamu berikan, YOLOv5 mendeteksi "person" atau orang dengan skor kepercayaan, yang terlihat di bagian kiri atas dari kotak pembatas (misalnya, "person 0.68" dan "person 0.79").

Berikut adalah penjelasan dari elemen-elemen utama dalam gambar berdasarkan YOLOv5:

Bounding Box (Kotak Biru):

Kotak biru di sekitar kepala dan sebagian tubuh menunjukkan area di mana YOLOv5 mendeteksi orang. Area ini mengelilingi objek yang terdeteksi untuk kejelasan visualisasi. Skor Kepercayaan:

Angka di samping label "person" (misalnya, 0.68 dan 0.79) menunjukkan tingkat keyakinan model dalam mengidentifikasi orang tersebut. Dalam gambar ini, YOLOv5 cukup yakin bahwa objek yang terdeteksi adalah orang, dengan skor kepercayaan 68Label Kelas:

Label kelas "person" menunjukkan bahwa YOLOv5 mengklasifikasikan objek yang terdeteksi sebagai orang, yang merupakan salah satu kategori yang sudah didefinisikan dalam dataset-nya. Kotak pembatas dan skor kepercayaan ini memberikan gambaran tentang seberapa baik kinerja model dalam mendeteksi objek. Untuk deteksi objek yang lebih presisi, kamu bisa melatih model dengan dataset khusus atau menyesuaikan parameter seperti ambang batas kepercayaan.

Jika menggunakan ini dalam proyek terkait NCT Dream atau video musik, YOLOv5 dapat membantu dalam melacak dan mengidentifikasi kehadiran orang (atau objek lainnya) dalam berbagai adegan atau frame video.




Figure 1: YOLOv5

```
from google.colab import files

# This will prompt you to upload a video file from your local system
uploaded = files.upload()

# To confirm the file is uploaded and check its name
for filename in uploaded.keys():
    print(f'Uploaded file: {filename}')
```

 Choose Files No file chosen Upload widget is only available when the cell has been executed

Saving inihaechan.mp4 to inihaechan.mp4
Uploaded file: inihaechan.mp4

Figure 2: Hasil Pelacakan YOLOv5