UNIVERSITAS
GADJAH MADA

Ekstraksi Data Tanda Vital dari Citra Monitor Fisiologis dengan YOLOv12 dan SVTRv2
Aizar Hafizh Soejadi, Afiahayati, S.Kom., M.Cs., Ph.D
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Ekstraksi Data Tanda Vital dari Citra Monitor Fisiologis dengan YOLOv12 dan SVTRv2

Oleh

Aizar Hafizh Soejadi

22/497615/PA/21435

Pemantauan tanda vital pasien merupakan bagian penting dalam lingkungan perawatan intensif. Namun, pencatatan nilai tanda vital dari monitor fisiologis masih dilakukan secara manual sehingga membutuhkan waktu dan rentan terhadap kesalahan pencatatan. Beberapa pendekatan OCR telah dikembangkan untuk mengotomatisasi proses tersebut, tetapi masih bergantung pada aturan pascapemrosesan yang kurang adaptif terhadap variasi tampilan monitor fisiologis. Penelitian ini bertujuan mengembangkan sistem ekstraksi data tanda vital dari citra monitor fisiologis melalui integrasi model YOLOv12 dan SVTRv2. YOLOv12 digunakan untuk melokalisasi nilai numerik sekaligus mengklasifikasikan parameter tanda vital, sedangkan SVTRv2 digunakan untuk mengenali karakter numerik pada *region of interest* hasil deteksi. Dataset penelitian terdiri atas 1.561 citra monitor fisiologis dengan 7.911 anotasi *bounding box* pada enam kelas tanda vital, yaitu HR, RR, SpO2, BP_SYS, BP_DIA, dan BP_MEAN. Evaluasi dilakukan terhadap model YOLOv12, model SVTRv2, dan sistem terintegrasi menggunakan metrik mAP, *word accuracy*, kelengkapan data, akurasi data, dan waktu inferensi. Hasil pengujian YOLOv12 menunjukkan bahwa seluruh varian mampu mendeteksi tanda vital dengan baik pada data uji reguler, dengan mAP_{50-95} tertinggi sebesar 0,963 pada YOLOv12s. Pada data uji khusus, YOLOv12x menunjukkan generalisasi terbaik dengan mAP_{50-95} sebesar 0,883. Sementara itu, SVTRv2-Base memperoleh *word accuracy* 99,96% pada 7.908 potongan citra yang dapat diverifikasi. Pada sistem terintegrasi, kombinasi YOLOv12x dan SVTRv2-Base menjadi kombinasi terbaik dengan kelengkapan data 90,18%, akurasi data 86,57%, dan rata-rata waktu inferensi 340,49 ms per citra. Hasil ini menunjukkan bahwa integrasi YOLOv12 dan SVTRv2 mampu mengekstraksi data tanda vital secara terstruktur dan cukup adaptif terhadap variasi tampilan monitor.

Kata kunci: monitor fisiologis, tanda vital, YOLOv12, SVTRv2, *optical character recognition*

UNIVERSITAS
GADJAH MADA

Ekstraksi Data Tanda Vital dari Citra Monitor Fisiologis dengan YOLOv12 dan SVTRv2
Aizar Hafizh Soejadi, Afiahayati, S.Kom., M.Cs., Ph.D
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Vital Sign Data Extraction from Physiological Monitor Image using YOLOv12 and SVTRv2

by

Aizar Hafizh Soejadi

22/497615/PA/21435

Vital sign monitoring is an important component in intensive care environments. However, vital sign values displayed on physiological monitors are still recorded manually, which is time-consuming and prone to recording errors. Several OCR-based approaches have been developed to automate this process, but they rely on postprocessing rules that are less adaptive to variations in physiological monitor. This study aims to develop a vital sign data extraction system from physiological monitor images through the integration of YOLOv12 and SVTRv2. YOLOv12 is used to localize numerical values and classify their corresponding vital signs, while SVTRv2 is used to recognize number within the detected regions of interest. The dataset consists of 1,561 physiological monitor images with 7,911 bounding box annotations across six vital sign classes, namely HR, RR, SpO2, BP_SYS, BP_DIA, and BP_MEAN. The YOLOv12 evaluation showed that all variants were able to detect vital signs effectively on regular test data, with the highest mAP50-95 of 0.963 achieved by YOLOv12s. On the special test data, YOLOv12x showed the best generalization performance with an mAP50-95 of 0.883. Meanwhile, SVTRv2-Base achieved a word accuracy of 99.96% on 7,908 verifiable cropped images. In the integrated system, the combination of YOLOv12x and SVTRv2-Base achieved the best performance, with 90.18% data completeness, 86.57% data accuracy, and an average inference time of 340.49 ms per image. These results indicate that the integration of YOLOv12 and SVTRv2 is capable of extracting vital sign data in a structured manner and is reasonably adaptive to variations in monitor displays.

***Keywords**: physiological monitor, vital sign, YOLOv12, SVTRv2, optical character recognition*