

Deteksi *Region of Interest* Tulang pada Citra *B-mode* secara Otomatis Menggunakan *Region Proposal Networks*

Tita Karlita^{1,4}, I Made Gede Sunarya¹, Joko Priambodo¹, Rika Rokhana¹, Eko Mulyanto Yuniarno^{2,3},
I Ketut Eddy Purnama^{2,3}, Mauridhi Hery Purnomo^{2,3}

Abstract— Bone imaging using ultrasound is a safe technique since it does not involve ionizing radiation and non-invasive. However, bone detection and localization to find its region of interest (RoI) is a challenging task because b-mode ultrasound images are characterized by high level of noise and reverberation artifacts. The image quality is user-dependent and the boundary between tissues is blurry, which makes it challenging to interpret images. In this paper, the deep learning approach using Region Proposal Networks was implemented to detect bone's RoI in b-mode images. The Faster Region-based Convolutional Neural Network model was fine-tuned to detect and determine the bone location in b-mode images automatically. To evaluate the results, in-vivo experiments were carried out using human arm specimens. A total of 1,066 b-mode bone images from six different subjects were used in the training phase and testing phase. The proposed method was successful in determining the bone RoI with the value of the mAP, the accuracy of detection, and the accuracy of localization of 0.87, 98.33%, and 95.99% respectively.

Intisari— Pencitraan tulang menggunakan *ultrasound* adalah sebuah teknik pencitraan yang aman karena bebas radiasi dan *non-invasive*. Namun, mendeteksi dan menentukan lokasi tulang untuk menemukan *region of interest* (RoI) adalah pekerjaan yang tidak mudah karena sifat citra *b-mode* yang memiliki derau yang tinggi dan *reverberation artifacts*. Kualitas citra dipengaruhi oleh pengguna dan batas antar area jaringan tidak jelas sehingga menyulitkan interpretasi. Dalam makalah ini, pendekatan *deep learning* menggunakan *Region Proposal Networks* diaplikasikan untuk mendeteksi RoI tulang dalam citra *b-mode* secara otomatis. Model arsitektur *Faster Region-based Convolutional Neural Network* disesuaikan agar dapat mendeteksi RoI tulang. Untuk

mengevaluasi hasil, eksperimen secara *in-vivo* dilakukan menggunakan spesimen lengan manusia. Sebanyak 1.066 citra *b-mode* tulang dari enam subjek yang berbeda digunakan dalam fase latih dan fase pengujian. Metode yang diusulkan berhasil menentukan RoI tulang dengan baik dengan nilai mAP, akurasi deteksi, dan akurasi penempatan RoI masing-masing sebesar 0,87, 98,33%, dan 95,99%.

Kata Kunci— *b-mode*, *deep learning*, deteksi, *faster R-CNN*, tulang, *region of interest*, *region proposal networks*, *ultrasound*.

I. PENDAHULUAN

Pendekatan *machine learning* di bidang pencitraan medis telah banyak diaplikasikan dalam berbagai macam kasus seperti deteksi kanker serviks [1], segmentasi citra sel tunggal *smear* serviks [2], deteksi tumor hati [3], klasifikasi citra termal kanker payudara [4], dan ortopedi [5]-[8]. Berbagai macam algoritme dan metode telah digunakan, di antaranya jaringan saraf tiruan, *graph-cut*, dan *fuzzy c-means*.

Di bidang ortopedi, aplikasi *ultrasound* digunakan untuk pemeriksaan patah tulang, pembuatan implan, atau untuk mengetahui dimensi dan geometri tulang dalam bentuk tiga dimensi. Penelitian pemanfaatan *ultrasound* di bidang ortopedi diterapkan dalam beberapa kasus berbeda. Pada umumnya metode yang dipakai adalah untuk membedakan fitur tulang dan bukan tulang menggunakan pendekatan heuristik. Karakteristik khusus yang digunakan adalah bahwa permukaan tulang tampak terang di citra *b-mode*. Kekurangan dari metode ini adalah kedalaman lokasi tulang harus sudah diketahui. Padahal ketika *probe ultrasound* digerakkan secara *freehand* dan bebas bergerak dalam enam derajat kebebasan, maka posisi tulang bisa di mana saja di dalam area gambar [9].

Dibandingkan dengan prosedur pencitraan yang lain seperti CT, MRI, dan *X-rays*, pencitraan menggunakan *ultrasound* lebih aman karena tidak memancarkan radiasi yang berbahaya dan tidak bersifat merusak. Namun, kualitas citra *b-mode* yang dihasilkan *ultrasound* sangat rendah karena keberadaan *speckle noise*, *reverberation*, *shadowing*, dan *signal dropout* [10]-[12]. Oleh karena itu, dibutuhkan usaha yang lebih untuk mendeteksi dan menentukan lokasi tulang dalam sebuah citra *b-mode ultrasound*.

Beberapa penelitian sebelumnya fokus ke kegiatan mengekstraksi kontur tulang, baik menggunakan pendekatan probabilistik, energetik, maupun regresi, dengan menggunakan fitur fase dan intensitas piksel [5]-[8]. Untuk segmentasi tulang pada citra *ultrasound*, beberapa pendekatan yang digunakan adalah *random forest classifier* [13], pemanfaatan fitur fase dan *shadowing* [14], penggunaan energi lokal dan *integrated back scattering* dengan pendekatan heuristik [15], kombinasi fitur probabilistik, *local binary pattern*, dan filter Gabor [16], penerapan model *snake* dan *tensor product B-splines*

¹Mahasiswa, Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember (ITS), Jln. Teknik Mesin, Kampus ITS, Sukolilo, Surabaya, Jawa Timur, 60111 Indonesia (telp: 031-5994251; fax: 031-5931237; e-mail: tita16@mhs.ee.its.ac.id, made.sunarya16@mhs.ee.its.ac.id, rika16@mhs.ee.its.ac.id)

²Dosen, Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember (ITS), Jln. Teknik Mesin, Kampus ITS, Sukolilo, Surabaya, Jawa Timur, 60111, Indonesia (telp: 031-5994251; fax: 031-5931237; e-mail: ekomulyanto@ee.its.ac.id, ketut@ee.its.ac.id, hery@ee.its.ac.id)

³Dosen, Departemen Teknik Komputer, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember (ITS), Jln. Teknik Mesin, Kampus ITS, Sukolilo, Surabaya, Jawa Timur, 60111, Indonesia (telp: 031-5922397; e-mail: ekomulyanto@ee.its.ac.id, ketut@ee.its.ac.id, hery@ee.its.ac.id)

⁴Dosen, Teknik Informatika, Departemen Teknik Informatika dan Komputer, Politeknik Elektronika Negeri Surabaya (PENS), Jln. Raya ITS, Kampus PENS, Sukolilo, Surabaya, Jawa Timur, 60111, Indonesia (telp: 031-5947280; fax: 031-5946114; email: tita@pens.ac.id)

approximation [17], dan pendekatan *dynamic programming* yang diaplikasikan pada beberapa fitur gambar. Di antara penelitian-penelitian tersebut, metode yang paling banyak digunakan adalah klasifikasi linier yang diterapkan pada beberapa kombinasi fitur nilai intensitas, *Laplacian of Gaussian filter*, dan *shadow effect*.

Penentuan lokasi tulang secara otomatis juga menjadi perhatian di bidang ortopedi, terutama untuk prosedur registrasi tulang menggunakan lebih dari satu jenis alat pencitraan tulang dan pada prosedur *computer-assisted orthopaedic surgery*. Sebuah penelitian menggunakan fitur frekuensi dan *Gaussian Mixture Models* untuk meregistrasi citra *ultrasound* dan CT untuk prosedur *computer-assisted orthopaedic surgery* [7], [18]. Penelitian lain mendeteksi penanda anatomi tulang paha menggunakan *convolutional neural network* (CNN), *shape statistics*, dan *graph cut optimization* [19].

Akhir-akhir ini, metode deteksi objek dan klasifikasi gambar berbasis *deep learning* terutama CNN terbukti sukses digunakan untuk pengenalan objek secara visual objek secara otomatis di berbagai bidang. Pada beberapa penelitian sebelumnya, metode *deep learning* telah diaplikasikan pada citra *ultrasound* berbagai organ seperti abdomen, jantung, dan tulang [20]-[23]. Dalam [20] digabungkan metode analisis tekstur tradisional dengan *deep learning* untuk deteksi otomatis organ abdominal. Hasil penentuan lokasi abdomen menggunakan pendekatan CNN dibandingkan dengan pendekatan analisis tradisional. *Deep belief networks* diaplikasikan untuk deteksi dan segmentasi *multi-atlas* pada ventrikel kiri jantung [21]. *Deep neural network* diaplikasikan untuk pelacakan lokasi ventrikel kiri pada kasus *endocardium* [22]. Sementara itu, CNN digunakan untuk pengenalan objek fetal abdomen, jantung, dan tulang [23].

Deteksi dan ekstraksi tulang secara manual dalam citra *b-mode* merupakan pekerjaan yang melelahkan dan membutuhkan usaha yang besar. Hasil deteksi tulang bervariasi bergantung dengan kemampuan dan kepakaran *observer* yang berbeda-beda [24]. Sementara itu, kinerja metode deteksi dan segmentasi tulang pada citra *b-mode* yang ada saat ini sensitif terhadap variasi data dan parameter *setting* [10]. Padahal, untuk mendeteksi dan menentukan lokasi tulang pada citra *b-mode* dibutuhkan metode yang cepat, akurat, dan otomatis.

Dalam makalah ini, *Region Proposal Networks* (RPN) digunakan untuk mengidentifikasi dan menentukan area-area dalam citra *b-mode* yang berisi objek tulang. Arsitektur *Faster Region-based Convolutional Neural Network* (*Faster R-CNN*) [25] digunakan dan dilatih untuk mendeteksi dan menentukan RoI tulang dalam citra *b-mode*. Arsitektur *Faster R-CNN* telah dilatih pada *ImageNet*. Selanjutnya dilakukan *transfer learning* menggunakan data citra *b-mode* tulang. Proses latih dan uji dilakukan di Google Colaboratory [26], [27] menggunakan GPU Tesla K80. Hasil deteksi dan penentuan lokasi menghasilkan RoI berupa *bounding box* area tulang dalam citra *b-mode*.

II. MATERIAL DAN METODE

Proses pendeteksian dan penentuan RoI tulang dalam citra *b-mode* terdiri atas empat langkah utama. Pertama, pengambilan dan penyiapan *dataset* citra *b-mode* tulang. Kedua, anotasi dan

melabeli objek tulang dalam citra *b-mode*. Ketiga, mengimplementasikan RPN menggunakan model arsitektur *Faster R-CNN* dengan memanfaatkan *framework TensorFlow Object Detection API*. Tahap akhir adalah mengevaluasi hasil deteksi dan penentuan RoI tulang dalam citra *b-mode*.

A. Dataset dan Pra-Proses Citra B-Mode

Sejumlah 1.066 citra *b-mode* tulang diambil dari enam subjek yang berbeda jenis kelamin dengan rentang usia antara 25 sampai 50 tahun. Data diambil menggunakan *freehand ultrasound* dua dimensi (2D), *TELEMED Linear Transducer L15-7L40H-5* dengan jangkauan frekuensi 7,0-15,0 MHz dan 39 mm *field of view*. Parameter *ultrasound* yang digunakan dalam pengambilan gambar meliputi *depth*, *dynamic range*, *power*, *gain*, dan frekuensi. Selama pengambilan data, nilai parameter ini diatur sedemikian rupa hingga didapatkan citra *b-mode* yang menggambarkan jaringan tulang beserta jaringan lunak di sekitarnya. Data masukan yang digunakan dalam sistem *deep learning* berupa *file* gambar berekstensi *.jpg*. Informasi lebih detail citra *b-mode* yang digunakan dalam studi ini diperlihatkan pada Tabel I. Dari semua citra, *dataset* dikelompokkan secara acak dengan proporsi 85% sebagai data latih dan sisanya sebagai data uji.

TABEL I
PROPORSI DATA LATIH DAN UJI CITRA B-MODE TULANG

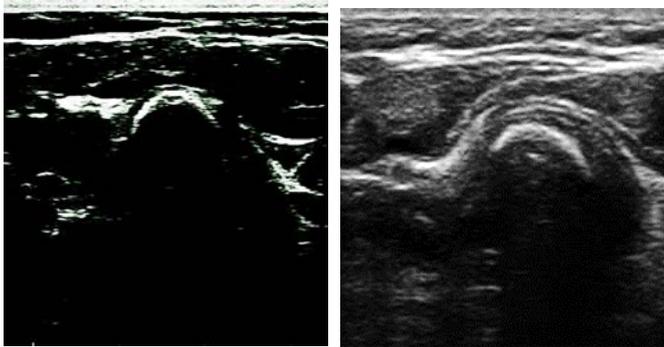
Spesimen	Jumlah Gambar		
	Latih	Uji	Jumlah
Subjek A (471x506)	264	29	293
Subjek B	117	29	146
Subjek C	232	29	261
Subjek D	303	29	332
Subjek Campuran (Subjek B, Subjek E, dan Subjek F)	0	34	34
Jumlah	916	150	1.066

Area pengambilan gambar tulang pada penelitian ini adalah lengan tangan bagian bawah, mulai dari tepat di bawah siku lengan sampai dengan tepat sebelum pergelangan tangan. Untuk data subjek A, B, C, dan D, pengambilan citra tulang dilakukan dengan cara menggerakkan *probe ultrasound* dengan kecepatan yang teratur. Posisi dan orientasi *probe ultrasound* diusahakan selalu tegak lurus terhadap permukaan lengan. Hasil pengambilan gambar untuk empat subjek tersebut disimpan dalam bentuk *file* video berekstensi *.avi*. Untuk selanjutnya, tiap *frame file* video diubah menjadi sejumlah *file* gambar. Hasil citra kelompok subjek ini menghasilkan citra *b-mode* dengan nilai kontras agak rendah dengan resolusi yang sama, yaitu 471x506 piksel.

Data subjek campuran diperoleh dengan cara mengambil data gambar berekstensi *.jpg* langsung dari tiga subjek yang berbeda (subjek B, E, dan F). Beberapa parameter *ultrasound* bernilai berbeda dengan kelompok subjek A, B, C dan D sebelumnya. Citra *b-mode* kelompok subjek ini memiliki tingkat kecerahan agak tinggi dan resolusi piksel yang bervariasi. Beberapa contoh citra *b-mode* tulang yang dihasilkan ditunjukkan pada Gbr. 1.

Karena spesimen tulang diambil dari tulang lengan bawah, yang disusun oleh dua tulang, yaitu *ulna* dan *radius*,

pengambilan tulang diusahakan terfokus pada salah satu tulang saja, sehingga secara umum satu citra *b-mode* berisi satu tulang. Namun, pada hasilnya ada beberapa citra yang mengandung sebagian permukaan tulang yang lain meskipun tidak tampak sempurna. Hal ini beberapa kali terjadi, terutama pada pengambilan *file* video (subjek A, B, C, dan D).



Gbr. 1 Contoh variasi citra *b-mode* tulang.

Data gambar mentah yang diperoleh dari perangkat lunak *Echowave TELEMED* tidak hanya berisi citra *b-mode*, tetapi juga mengandung informasi parameter *setting ultrasound* dan *frame*. Karena data ini tidak diperlukan dalam proses deteksi dan penentuan RoI tulang, maka informasi ini dibuang dengan cara memotong gambar, sehingga data masukan fokus hanya berupa citra *b-mode*.

B. Anotasi Data

Anotasi data dilakukan dengan cara melabeli area tulang sesuai arahan pakar di bidang radiologi muskuloskeletal. Dalam citra *b-mode*, tulang memiliki area dengan tingkat kecerahan yang lebih tinggi dari jaringan lunak di sekelilingnya. Karena sinyal *ultrasound* tidak dapat menembus jaringan padat, maka pada area citra di bawah tulang terdapat *shadow* yang menyebabkan area tersebut memiliki tingkat kecerahan rendah dengan warna mendekati hitam [10].

Pelabelan dilakukan dengan cara *tagging* secara manual pada semua gambar, baik pada data latih maupun data uji. Anotasi yang dihasilkan dari proses ini berupa *bounding box* berbentuk bidang persegi empat dengan keempat sisi berada tepat di sekitar luar area tulang. Informasi yang diperoleh dari hasil pelabelan tulang berupa koordinat pojok kiri atas *bounding box* beserta dimensi panjang dan lebar, nama kelas objek, dan *file* gambar yang berkaitan disimpan sebagai *file* berekstensi *.xml* dalam format *PASCAL VOC*.

Selanjutnya, agar hasil anotasi siap dipakai sebagai masukan *TensorFlow Object Detection API*, maka hasil anotasi diubah menjadi format *file TFRecord* dengan ekstensi *.record*. Dalam *file TFRecords*, citra *b-mode* diubah menjadi *array NumPy*. *TensorFlow* menggunakan format *Google protobuf text* berupa *file* berekstensi *.pbtxt* untuk menyimpan informasi label kelas.

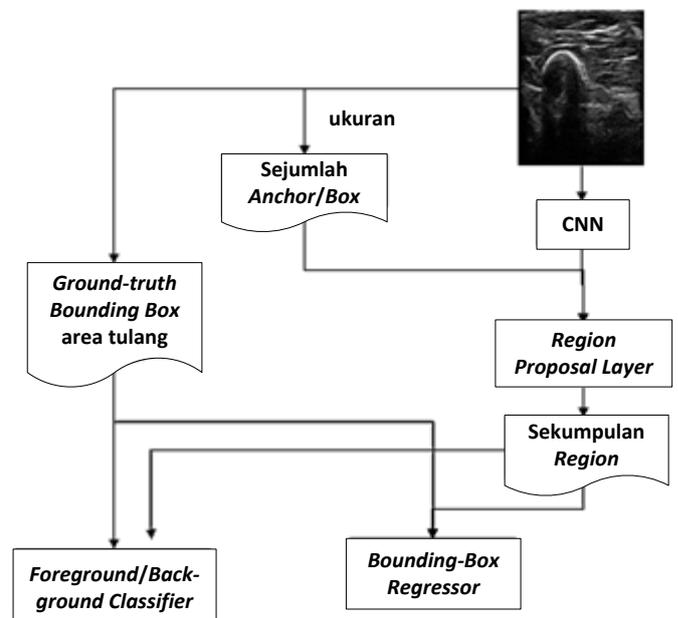
C. Region Proposal Networks (RPN)

Dalam makalah ini, RPN digunakan untuk menemukan sejumlah area berupa *bounding box* yang berkemungkinan berisi objek tulang dalam citra *b-mode*. Model arsitektur *Faster R-CNN* [25] digunakan untuk mendeteksi dan menentukan

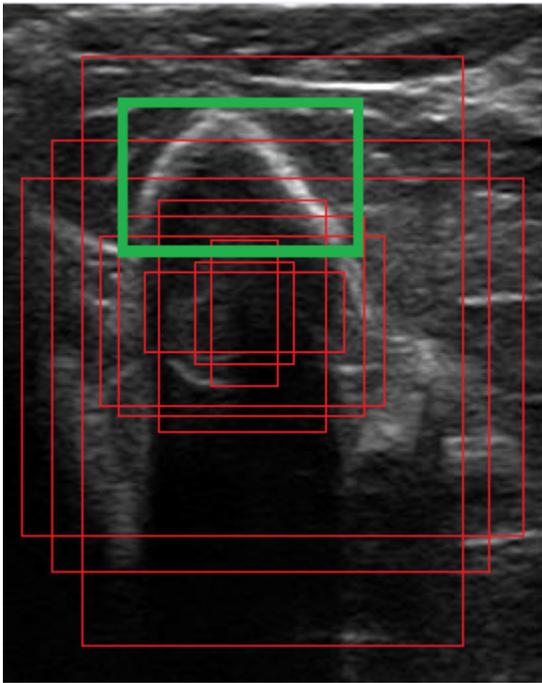
lokasi tulang. *Faster R-CNN* adalah model deteksi objek yang telah mengalami *pre-training* pada *ImageNet*. Selanjutnya dilakukan *tuning* untuk data citra *b-mode* tulang dengan satu kelas *bounding box*. *Faster R-CNN* merupakan sistem pendeteksi objek berbasis CNN yang disusun oleh dua buah modul utama [25], [28]. Modul pertama disebut dengan RPN dan modul kedua adalah *Fast R-CNN detector*. Sistem yang dibentuk oleh dua modul ini secara keseluruhan disebut dengan *unified network*.

Dalam makalah ini, RPN berupa *deep fully convolutional network* yang bertugas untuk menentukan area-area dari citra *b-mode* dengan berbagai ukuran yang disebut dengan *region proposal*. RPN menentukan dan memprediksi tingkat keberadaan objek tulang (*objectness score*) pada tiap *region proposal*, sehingga keluaran RPN adalah sejumlah *region proposal* beserta *objectness score* masing-masing yang disebut dengan *anchor*. *Anchor* yang berisi gambar tulang disebut *foreground* dan sebaliknya disebut *background*. RPN menentukan satu *anchor* yang berkemungkinan besar sebagai *foreground*, yaitu *bounding box* yang berisi objek tulang. Gbr. 2 menggambarkan arsitektur RPN saat fase latih.

Sebuah *anchor* direpresentasikan dalam bentuk sebuah *box*. Gambaran letak sembilan buah *anchor* di posisi (320,320) dalam sebuah citra dengan ukuran 800x600 diilustrasikan dalam Gbr. 3. *Anchor* berwarna merah dengan garis tipis merepresentasikan tiga skala atau ukuran berbeda: 128x128, 256x256, dan 512x512. Masing-masing warna memiliki *box/anchor* dengan rasio tinggi dan lebar sebesar 1:1, 1:2, dan 2:1. *Anchor* berwarna hijau dengan garis tebal adalah *ground truth box*. Setelah mengaplikasikan CNN, dihasilkan sejumlah *feature map*. Tiap posisi dalam *feature map* memiliki sembilan *anchor* dan tiap *anchor* memiliki dua kemungkinan label, yaitu *background* atau *foreground*.



Gbr. 2 Arsitektur *Region Proposal Network*.



Gbr. 3 Ilustrasi lokasi sejumlah *anchor box* dalam sebuah citra *b-mode*.

Pada fase latih, sebuah *classifier* dibuat dengan terlebih dahulu membuat *dataset* latih. Data latih terdiri atas sekumpulan *anchor* dan sekumpulan *ground-truth*, yaitu citra *b-mode* yang telah dianotasi dan *bounding box* yang telah dilabeli dengan kelas tulang. Selanjutnya adalah melabeli sekumpulan *anchor* dengan memanfaatkan sekumpulan *ground-truth box*. Pelabelan *anchor-anchor* dilakukan dengan kriteria bahwa sekumpulan *anchor* yang memiliki *overlap* (IoU) yang tinggi dengan *ground-truth* didefinisikan sebagai *foreground*, dan sebaliknya, nilai *overlap* (IoU) yang rendah didefinisikan sebagai *background*. IoU *threshold* dalam penelitian ini adalah 0,6. Sampai dengan tahap ini, semua *anchor* telah terlabeli.

Sekumpulan *anchor* yang memiliki label *background* tidak diproses oleh *regressor* karena tidak tersedia *ground-truth* untuk *anchor-anchor* tersebut. Di makalah ini, *Regressor Loss Function* menggunakan *smooth-L1 loss* pada posisi pojok kiri atas *box* (x,y) dan logaritma dari ketinggian dan dalaman *box*. Secara keseluruhan, RPN adalah kombinasi dari *classification loss* dan *regression loss* (1).

$$L_{loc}(t^v, v) = \sum_{i \in \{x,y,w,h\}}^{n} \text{smooth}_{L_1}(t_i^u - v_i) \quad (1)$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{sebaliknya} \end{cases} \quad (2)$$

Setelah proses RPN bekerja, maka didapatkan sekumpulan *proposed region* dengan ukuran yang berbeda-beda. Ukuran *region* yang berbeda-beda ini berarti *CNN feature map* juga berbeda-beda ukurannya. Struktur yang bekerja secara efisien sulit didapat untuk fitur dengan ukuran berbeda. *Region of Interest Pooling* (*RoI Pooling*) menyelesaikan permasalahan ini dengan mereduksi *feature map* ke dalam ukuran yang sama.

Tidak seperti *Max-Pooling* yang berukuran tetap, *RoI Pooling* membagi *feature map* masukan menjadi sejumlah k buah, yaitu *region-region* yang memiliki ukuran yang hampir sama. Selanjutnya adalah mengaplikasikan *Max-Pooling* pada tiap *region*. Oleh karena itu, keluaran *RoI Pooling* selalu berjumlah k buah tanpa mengindahkan berapa pun ukuran masukan. Dalam *RoI Pooling layer*, object proposal mengekstraksi area objek dari sejumlah *feature map*. Fitur-fitur yang terekstrak diaplikasikan ke semua *feature extractor* untuk mendapatkan probabilitas prediksi kelas dan *bounding box* untuk tiap *region* proposal. Langkah terakhir adalah mengklasifikasikan objek-objek yang ditemukan di gambar beserta lokasinya dalam bentuk *bounding box*.

D. Implementasi Deep Learning dengan Region Proposal Networks

Sistem deteksi dan penentuan RoI tulang dibuat menggunakan *framework TensorFlow Object Detection API* yang dibangun di Google Colaboratory [26]. *Colaboratory notebooks* berbasis *Jupyter Notebook*. Proses *tuning* dan tes dijalankan di komputer dengan spesifikasi Ubuntu 17.10 64 bit yang didukung dengan prosesor Intel(R) Xeon(R), *two cores @2.3 GHz* dan RAM 13 GB, dilengkapi dengan NVIDIA GPU Tesla K80 (GK210 chipset), ~12 GB RAM, 2496 CUDA *cores @560 MHz* [27].

Dalam makalah ini, *pre-trained model* yang digunakan adalah *faster_rcnn_inception_v2_coco_2018_01_28* yang telah dilatih sebelumnya pada *dataset Common Objects in Context* (COCO). *Pipeline default* digunakan dan penyesuaian konfigurasi dilakukan pada fase *tuning* agar dapat mengenali objek tulang dalam citra *b-mode*.

E. Metrik Evaluasi

Metrik yang digunakan untuk mengevaluasi kinerja metode dalam makalah ini adalah *precision*, *mean average precision* (mAP), *recall*, dan waktu komputasi. *Precision* menggambarkan tingkat keyakinan hasil deteksi. *Precision* merupakan ukuran keakuratan deteksi yang dilakukan, yaitu persentase deteksi positif (TP) adalah benar yang dirumuskan dengan (3), dengan TP adalah *true positive*, dan FP adalah *false positive*. *Recall* menggambarkan persentase keyakinan objek yang terdeteksi oleh *detector*. *Recall* merupakan ukuran tingkat kebaikan prediksi positif (TP) sebagaimana dirumuskan dengan (4), dengan FN adalah *false negative*. Secara umum, semakin tinggi nilai *precision*, maka nilai *recall* semakin rendah, dan sebaliknya. Hasil deteksi tulang disebut TP jika IoU melebihi 60%, sebagaimana dirumuskan pada (5). Dalam (5), *true bounding box* dinotasikan sebagai *tbb*, sedangkan *predicted bounding box* dinotasikan sebagai *pbb*, sementara IoU adalah ukuran banyaknya *overlap* antara area *bounding box* yang ditemukan dan *groundtruth*.

$$\text{precision} = \frac{TP}{(TP + FP)} \quad (3)$$

$$\text{recall} = \frac{TP}{(TP + FN)} \quad (4)$$

$$\text{IoU} = \frac{\text{area}(tbb \cap pbb)}{\text{area}(tbb \cup pbb)} > 0,6. \quad (5)$$

Dalam makalah ini, sebuah tulang hanya dapat memiliki satu *bounding box*, sehingga jika terdapat beberapa *bounding box* untuk sebuah tulang, maka hanya satu yang dianggap TP dan yang lain dianggap FP. Objek tulang yang tidak terdeteksi *bounding box* disebut sebagai FN. mAP menyajikan informasi *precision* dan *recall* dalam satu angka. Kurva *precision-recall* dibentuk dengan mengurutkan semua *bounding box* yang terprediksi yang ditetapkan (*assigned*) ke kelas tulang berdasarkan *confidence rate*. Selanjutnya untuk tiap prediksi, nilai *recall* dan nilai *precision* dihitung. *Recall* didefinisikan sebagai proporsi TP dalam nilai *threshold* tertentu. Sedangkan waktu komputasi adalah waktu yang dibutuhkan oleh sistem untuk melakukan proses latih.

Evaluasi visual juga dilakukan untuk mengukur kinerja sistem. Variabel yang diukur adalah rasio keberhasilan sistem dalam mendeteksi tulang (akurasi deteksi) dan ketepatan lokasi *bounding box* pada fitur tulang secara visual (akurasi *bounding box*). Akurasi deteksi diukur dengan membandingkan jumlah tulang yang terdeteksi (satu tulang utama per citra *b-mode*) dan total jumlah *b-mode* yang diuji, seperti pada (6). Akurasi *bounding box* diukur dengan cara membandingkan jumlah *b-mode* dengan lokasi *bounding box* yang tepat keempat sisinya pada fitur tulang (BB) dengan total jumlah *b-mode* yang terdeteksi, seperti pada (7).

$$\text{Akurasi deteksi} = \left(\frac{TP}{N}\right) \times 100\% \quad (6)$$

$$\text{Akurasi bounding box} = \left(\frac{BB}{TP}\right) \times 100\%. \quad (7)$$

III. HASIL EKSPERIMEN DAN DISKUSI

Eksperimen didasarkan atas dua skenario yang dibuat, yaitu untuk mengevaluasi hasil *tuning* dan keakuratan deteksi dan penentuan tulang dalam citra *b-mode*, serta evaluasi hasil tes secara visual.

A. Skenario Eksperimen

Untuk mengukur kinerja metode, dibuat dua buah skenario dan dua buah kelompok *dataset*. Pengelompokan *dataset* ditunjukkan pada Tabel II. *Dataset* kelompok 1 menggunakan data latih sebanyak 264 gambar dari satu subjek saja, yaitu subjek A. Pada fase uji digunakan variasi data dari banyak subjek (subjek A, B, C, dan D) dan data citra *b-mode* yang memiliki nilai kecerahan dan resolusi yang berbeda (subjek campuran). *Dataset* kelompok 2 menggunakan data latih sebanyak 916 gambar dari empat subjek berbeda, yaitu subjek A, B, C, dan D. Pada fase uji digunakan variasi data dari banyak subjek (subjek A, B, C, dan D) dan data citra *b-mode* yang memiliki nilai kecerahan dan resolusi yang berbeda dengan data latih (subjek campuran).

Skenario pertama bertujuan untuk mengukur efek keragaman citra *b-mode* dan jumlah data latih terhadap hasil deteksi dan penentuan lokasi tulang. Untuk itu, proses latih dijalankan pada *dataset* kelompok 1 dan kelompok 2 masing-masing satu kali *running* dengan jumlah iterasi dan *learning rate* yang sama, yaitu iterasi sebanyak 10.000 kali dan *learning rate* sebesar 0,0002. Skenario kedua bertujuan untuk mengukur efek penentuan *learning rate* terhadap hasil deteksi dan penentuan lokasi tulang. Untuk tujuan ini, proses latih

dijalankan pada *dataset* kelompok 2 saja sebanyak dua kali *running*, masing-masing dengan jumlah iterasi yang sama (3.000 iterasi) dan *learning rate* yang berbeda (0,0002 dan 0,002).

TABEL II
PEMBAGIAN DATASET UNTUK SKENARIO UJICOBA

Kelompok	Jumlah Gambar	
	Latih	Uji
Kelompok 1	Subjek A (264)	Subjek A (29)
		Subjek B (29)
		Subjek C (29)
		Subjek D (29)
		Subjek Campuran (34) (Subjek B, E, dan F)
Kelompok 2	Subjek A, B, C, dan D (916)	Subjek A (29)
		Subjek B (29)
		Subjek C (29)
		Subjek D (29)
		Subjek Campuran (34) (Subjek B, E, dan F)

Ket.: Angka dalam tanda kurung menunjukkan jumlah data gambar.

B. Efek Variasi dan Jumlah Data Latih

Pada skenario eksperimen pertama, proses latih dijalankan dengan iterasi sebanyak 10.000 kali dan *learning rate* sebesar 0,0002 untuk semua kelompok data. Grafik mAP, *recall*, dan *loss* masing-masing diperlihatkan pada Gbr. 4.

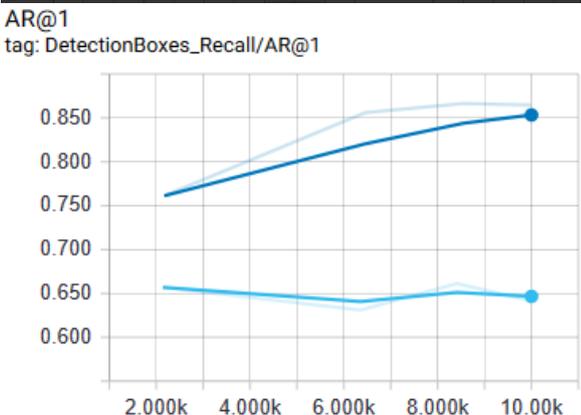
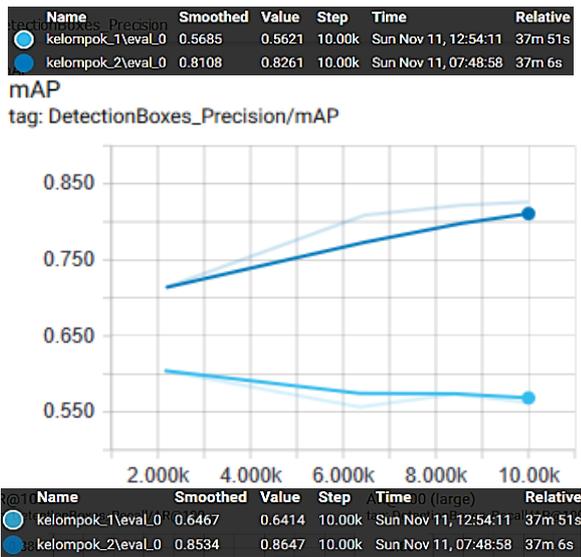
Gbr. 4 menampilkan sejumlah grafik yang menunjukkan perbandingan hasil proses *tuning* sehubungan dengan pengaruh jumlah dan variasi data latih terhadap model yang dihasilkan. Berdasarkan perbandingan mAP dan *recall*, data latih yang berjumlah banyak dan bervariasi tinggi memiliki nilai mAP dan *recall* yang cenderung semakin tinggi seiring dengan bertambahnya jumlah iterasi, tetapi berlaku sebaliknya untuk data latih yang berjumlah sedikit dan variasi yang kecil. Data latih dengan jumlah dan variasi tinggi memberikan hasil yang lebih bagus dengan perbedaan nilai mAP dan *recall* pada akhir iterasi masing-masing sebesar 0,264 dan 0,2233.

Tampak pada grafik *loss_1* bahwa nilai *loss* untuk kedua grafik sama-sama turun secara signifikan saat awal iterasi sampai dengan sekitar seribu kali iterasi. Di atas iterasi seribu kali, nilai *loss* secara umum mulai stabil, tetapi untuk data latih dengan jumlah dan variasi tinggi terdapat beberapa fluktuasi nilai sampai dengan akhir iterasi. Perbedaan nilai *loss* di akhir iterasi sebesar 0,0162. Dari sejumlah grafik ini dapat disimpulkan bahwa data latih yang memiliki variasi dan jumlah data yang tinggi secara umum menghasilkan model dan akurasi yang lebih baik.

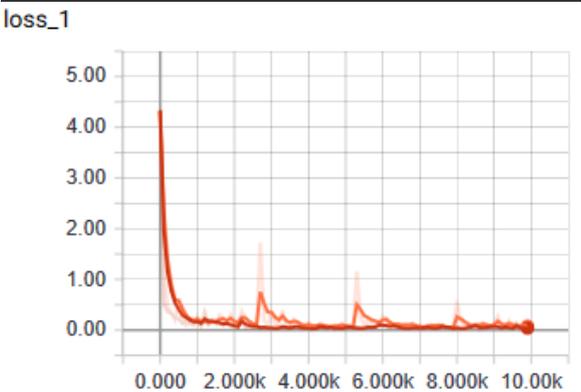
Hasil akurasi penempatan *bounding box* dan akurasi deteksi dan penentuan lokasi tulang pada citra *b-mode* tampak seperti pada Tabel III. Tampak pada tabel tersebut bahwa data latih yang memiliki variasi dan jumlah data yang tinggi secara umum menghasilkan akurasi yang lebih unggul, yaitu 100%.

C. Efek Penentuan Learning Rate

Berdasarkan eksperimen skenario pertama, nilai iterasi 3.000 dipilih pada eksperimen kedua ini karena pada level tersebut model yang terbentuk sudah stabil. Grafik mAP, *recall*,

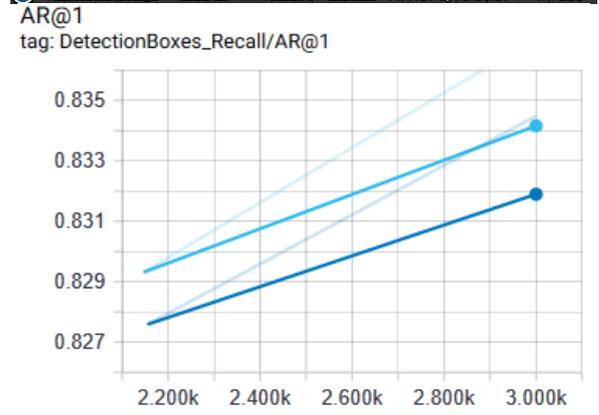
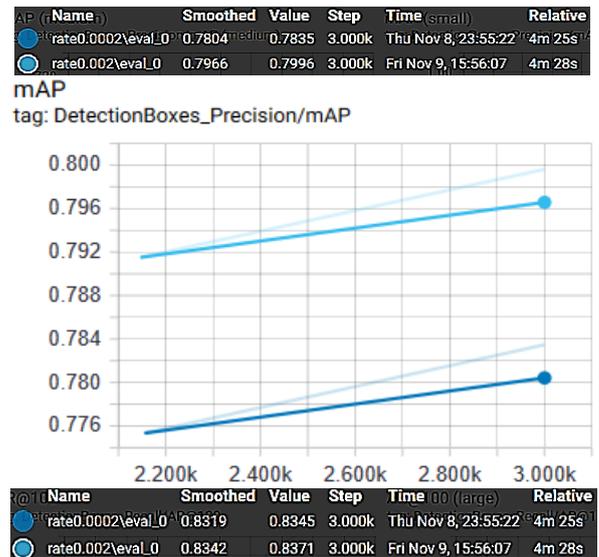


Name	Smoothed	Value	Step	Time	Relative
kelompok_1\	0.04115	0.03380	9.901k	Sun Nov 11, 12:53:09	47m 18s
kelompok_2\	0.07966	0.05000	9.901k	Sun Nov 11, 07:47:57	46m 32s

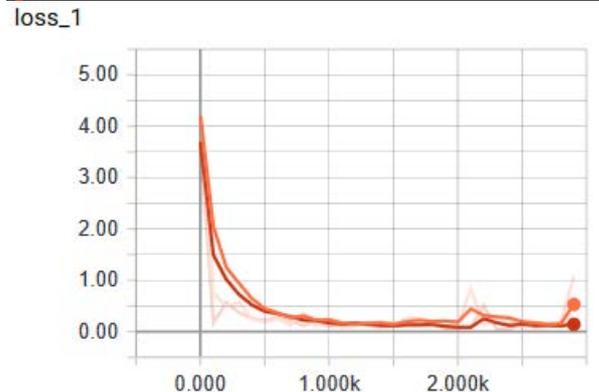


Gbr. 4 Perbandingan nilai mAP, recall per satu deteksi bounding box, dan loss pada eksperimen skenario pertama (mengevaluasi efek variasi dan jumlah data latih, *smoothing* = 0,6).

dan *loss* dari hasil menjalankan sistem untuk skenario eksperimen kedua masing-masing ditunjukkan pada Gbr. 5. Proses *tuning* hanya melibatkan *dataset* kelompok 2. Gbr. 5 Menampilkan beberapa grafik yang menunjukkan perbandingan hasil proses *tuning* sehubungan dengan pengaruh penentuan nilai *learning rate* terhadap model yang dihasilkan.



Name	Smoothed	Value	Step	Time	Relative
rate0.0002\	0.5307	1.092	2.901k	Thu Nov 8, 23:54:21	13m 54s
rate0.002\	0.1443	0.1903	2.901k	Fri Nov 9, 15:55:05	13m 55s



Gbr. 5 Perbandingan nilai mAP, recall per satu deteksi bounding box, dan loss pada eksperimen skenario kedua (mengevaluasi efek nilai *learning rate*, *smoothing* = 0,6).

Berdasarkan perbandingan mAP dan *recall*, secara umum semakin tinggi jumlah iterasi, semakin tinggi juga nilai mAP dan *recall* yang dihasilkan. Namun, *learning rate* 0,002 memberikan hasil yang sedikit lebih bagus dengan perbedaan nilai mAP dan *recall* pada iterasi ke 3.000 masing-masing sebesar 0,0161 dan 0,0026.

TABEL III
AKURASI PENEMPATAN *BOUNDING BOX* DAN AKURASI DETEKSI PADA EKSPERIMEN SKENARIO PERTAMA

	Nama dataset	Akurasi bounding box (%)	Akurasi deteksi (%)
Kelompok 1	Subjek A	100	100
	Subjek B	59,09	75,86
	Subjek C	85,71	96,55
	Subjek D	100	100
	Subjek Campuran	75,00	94,12
	Rata-rata	83,96	93,31
Kelompok 2	Subjek A	100	100
	Subjek B	100	100
	Subjek C	100	100
	Subjek D	100	100
	Subjek Campuran	100	100
	Rata-rata	100	100

TABEL IV
AKURASI PENEMPATAN *BOUNDING BOX* DAN AKURASI DETEKSI PADA EKSPERIMEN SKENARIO KEDUA

	Nama dataset	Akurasi bounding box (%)	Akurasi deteksi (%)
LR=0,002	Subjek A	100	100
	Subjek B	100	100
	Subjek C	100	100
	Subjek D	100	100
	Subjek Campuran	100	100
	Rata-rata	100	100
LR= 0,0002	Subjek A	100	100
	Subjek B	100	100
	Subjek C	100	100
	Subjek D	100	100
	Subjek Campuran	100	100
	Rata-rata	100	100

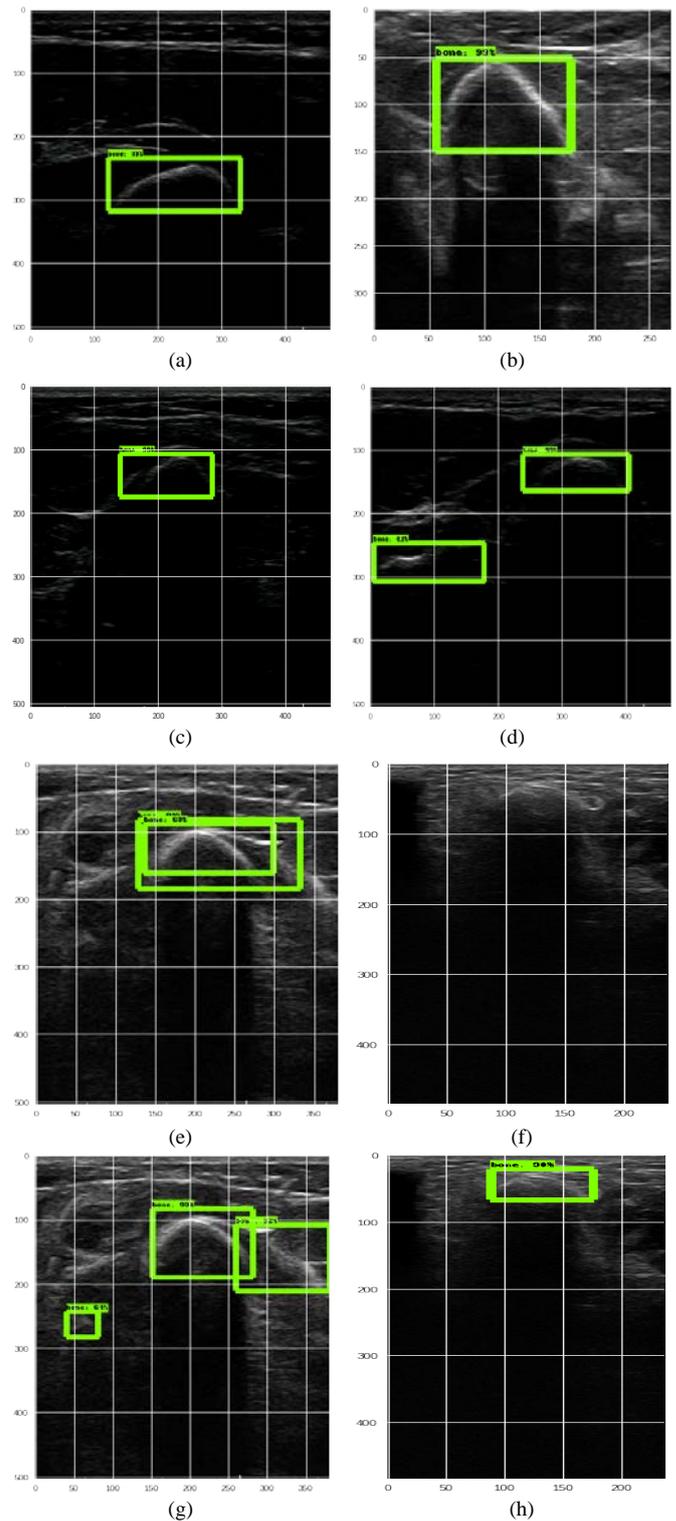
TABEL V
PERBANDINGAN WAKTU KOMPUTASI, MAP, DAN RECALL UNTUK SEMUA SKENARIO EKSPERIMEN

	Waktu (menit,detik)	mAP (IoU=,50:,95)	mAP (IoU=,75)	Recall (I)
A	48:40	0,60	0,68	0,66
B	48:19	0,71	0,87	0,76
C	15:37	0,78	0,96	0,83
D	15:43	0,79	0,99	0,83
Rata-rata		0,72	0,87	0,77

Keterangan:

- A=skenario 1, kelompok 1, iterasi 10.000, *learning rate* 0,0002
- B=skenario 1, kelompok 2, iterasi 10.000, *learning rate* 0,0002
- C=skenario 2, kelompok 2, iterasi 3.000, *learning rate* 0,0002
- D=skenario 2, kelompok 2, iterasi 3.000, *learning rate* 0,002.

Tampak pada grafik *loss_1*, bahwa nilai *loss* untuk kedua grafik sama-sama turun secara signifikan saat iterasi baru dimulai sampai dengan seribu kali iterasi. Di atas iterasi seribu kali, nilai *loss* secara umum mulai stabil, tetapi nilai *loss* untuk percobaan dengan penentuan *learning rate* 0,0002 sempat mengalami kenaikan pada akhir iterasi. Perbedaan nilai *loss* di akhir iterasi sebesar 0,9017. Dari grafik ini dapat disimpulkan bahwa nilai *learning rate* 0,002 secara umum menghasilkan model dan akurasi yang lebih baik.



Gbr. 6 Sejumlah hasil deteksi dan penentuan RoI tulang pada citra *b-mode* untuk berbagai macam skenario eksperimen, jumlah data, jumlah iterasi, dan *learning rate*.

Hasil evaluasi secara visual untuk akurasi penempatan *bounding box* dan akurasi deteksi dan penentuan RoI tulang pada citra *b-mode* disajikan di Tabel IV. Tampak pada tabel bahwa penentuan *learning rate* pada eksperimen ini tidak

signifikan memengaruhi akurasi nilai keduanya. Data latih dengan jumlah dan variasi tinggi dengan nilai *rate* 0,002 atau 0,0002, keduanya memberikan akurasi *bounding box* dan akurasi deteksi masing-masing sebesar 100%

Untuk semua skenario eksperimen, waktu komputasi, mAP dan *recall* disajikan dalam Tabel V. Dibutuhkan waktu *tuning* sekitar lima belas menit untuk melatih kurang lebih 264 gambar. Waktu komputasi naik menjadi hampir empat kali lipat untuk melatih sekitar 916 gambar. Nilai mAP dan *recall* tertinggi dimiliki oleh data latih dengan jumlah dan variasi tertinggi dengan jumlah iterasi 3.000 kali dan *learning rate* sebesar 0,002 (D).

D. Hasil Deteksi dan Penentuan RoI secara Visual

Hasil eksperimen terhadap semua citra latih menghasilkan berbagai macam deteksi dan penentuan RoI. Keragaman hasil deteksi dan penentuan lokasi RoI secara visual ditunjukkan pada Gbr. 6.

Gbr. 6(a) dan Gbr. 6(b) adalah hasil deteksi dan penentuan RoI dengan keempat posisi *bounding box* tepat berada di sekitar objek tulang. Tampak bahwa sistem mampu mendeteksi dengan baik meskipun bentuk dan tingkat kecerahan serta resolusi citra *b-mode* berbeda. Gbr. 6(c) adalah hasil deteksi dan penentuan RoI yang kurang tepat, karena batas *bounding box* pada satu atau lebih sisinya tidak tepat berada di sekitar objek tulang. Hal ini kemungkinan dapat terjadi karena adanya refleksi sinyal *ultrasound* sehingga jaringan lunak di atas tulang memiliki kecerahan dan penampilan yang mirip dengan tulang.

Gbr. 6(d) dan Gbr. 6(g) adalah hasil pendeteksian RoI tulang yang berjumlah lebih dari satu. Hal ini dapat terjadi ketika dalam satu citra *b-mode* terkandung lebih dari satu tulang. Jika diamati, tampak bahwa tulang utama diidentifikasi dengan nilai IoU sebesar 99% (TP), sementara tulang yang tidak tampak sempurna diidentifikasi dengan nilai IoU sebesar 81% (FP). Demikian juga untuk Gbr. 6(e) dan Gbr. 6(g), terdapat lebih dari satu *bounding box* dalam satu gambar. Kesalahan deteksi *bounding box* ini terjadi karena refleksi yang menyebabkan jaringan lunak di atas tulang memiliki tingkat kecerahan mirip tulang, sehingga menghasilkan *bounding box* lebih dari satu dengan beragam nilai IoU. Dan sekali lagi, tulang utama teridentifikasi dengan nilai IoU sebesar 99%. Dalam hal ini, tulang dengan *bounding box* tertinggi, yaitu bernilai IoU 99%, dianggap sebagai TP dan sisanya sebagai FP.

Gbr. 6(f) dan Gbr. 6(h) adalah citra *b-mode* yang sama dan *bounding box* dihasilkan oleh eksperimen yang berbeda. Objek tulang dalam gambar terlihat kabur. Ketika jumlah data latih dan variasi citra tinggi, RoI berupa *bounding box* tersebut dapat terdeteksi (TP), baik menggunakan *learning rate* 0,002 maupun 0,0002. Dan sebaliknya, objek tulang tidak terdeteksi (FN) untuk data latih dengan variasi dan jumlah data yang rendah.

IV. KESIMPULAN

Dalam makalah ini disajikan penggunaan metode *deep learning* dengan memanfaatkan RPN dengan model arsitektur *Faster R-CNN* untuk mendeteksi dan menentukan lokasi RoI tulang pada citra *b-mode*. Uji coba dengan menggunakan data dengan karakteristik citra *b-mode* yang bervariasi dari enam

subjek berbeda membuktikan bahwa sistem mampu mendeteksi dan menentukan RoI tulang dengan cukup akurat. Evaluasi hasil deteksi dan penentuan RoI menghasilkan nilai mAP, akurasi deteksi, dan akurasi penempatan lokasi masing-masing sebesar sebesar 0,87, 98,33%, dan 95,99%. Waktu yang dibutuhkan pada fase *tuning* dengan menggunakan Google Colaboratory cukup cepat dengan rata-rata waktu komputasi 15 menit per sekitar 250 citra *b-mode*.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Kementerian Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia dan Kementerian Keuangan Republik Indonesia, serta Lembaga Pengelola Dana Pendidikan (LPDP) yang telah mendukung penelitian ini melalui dana hibah BUDI DN-2016.

REFERENSI

- [1] D.A. Dharmawan, "Deteksi Kanker Serviks Otomatis Berbasis Jaringan Saraf Tiruan LVQ dan DCT," *J. Nas. Tek. Elektro dan Teknol. Inf.* Vol. 3, No. 4, hal. 269–272, 2014.
- [2] N.P. Husain dan C. Fatchah, "Segmentasi Citra Sel Tunggal Smear Serviks Menggunakan Radiating Component Normalized Generalized GVFS," *J. Nas. Tek. Elektro dan Teknol. Inf.*, Vol. 6, No. 1, hal. 107–114, 2017.
- [3] N. Syakrani, Y. Widhiyana, dan A.A. Efendi, "Deteksi Tumor Hati dengan Graph Cut dan Taksiran Volume Tumornya," *J. Nas. Tek. Elektro dan Teknol. Inf.*, Vol. 7, No. 1, hal. 35–43, 2018.
- [4] O. Herliana, T.S. Widodo, dan I. Soesanti, "Klasifikasi Nonsupervised Citra Thermal Kanker Payudara Berbasis Fuzzy C-MEANS," *J. Nas. Tek. Elektro dan Teknol. Inf.*, Vol. 1, No. 3, hal. 55–59, 2012.
- [5] T. Karlita, E.M. Yuniarno, I.K.E. Purnama, dan M.H. Purnomo, "Automatic Bone Outer Contour Extraction from B-Modes Ultrasound Images Based on Local Phase Symmetry and Quadratic Polynomial," *Second Int. Work. Pattern Recognit. (IWPR 2017)* 2017, pp. 165–170.
- [6] P.J.S. Gonçalves dan P. Torres, "Extracting Bone Contours in Ultrasound Images: Energetic Versus Probabilistic Methods," *Rom. Rev. Precis. Mech. Opt. Mechatronics*. Vol. 20, No. 37, hal. 105–110, 2010.
- [7] I. Hacıhaliloglu, P. Guy, A.J. Hodgson, dan R. Abugharbieh, "Automatic Extraction of Bone Surfaces from 3D Ultrasound Images in Orthopaedic Trauma Cases," *Int. J. Comput. Assist. Radiol. Surg.*, Vol. 10, hal. 1279–1287, 2015.
- [8] J. Kowal, C. Amstutz, F. Langlotz, H. Talib, dan M.G. Ballester, "Automated Bone Contour Detection in Ultrasound B-Mode Images For Minimally Invasive Registration in Computer-Assisted Surgery – An In Vitro Evaluation," *Int. J. Med. Robot. Comput. Assist. Surg. MRCAS*, Vol. 3, No. 4, hal. 341–348, 2007.
- [9] R.W. Prager, R.N. Rohling, A.H. Gee, dan L. Berman, "Rapid Calibration for 3-D Freehand Ultrasound," *Ultrasound Med. Biol.* Vol. 24, No. 6, hal. 855–869, 1998.
- [10] A.K. Jain dan R.H. Taylor, "Understanding Bone Responses in B-Mode Ultrasound Images and Automatic Bone Surface Extraction Using a Bayesian Probabilistic Framework," *Proc. SPIE, Med. Imaging 2004 Ultrason. Imaging Signal Process.*, 2004, Vol. 5373, hal. 131–142.
- [11] V. Chan dan A. Perlas, "Basics of Ultrasound Imaging," in *Atlas Ultrasound-Guided Proced. Interv. Pain Manag.*, S.N. Narouze, Ed., Toronto, ON, Canada, Springer Science+Business Media, 2011, hal. 13–20.
- [12] K.E. Purnama, M.H.F. Wilkinson, A.G. Veldhuizen, P.M.A. Van Ooijen, J. Lubbers, J.G.M. Burgerhof, T.A. Sardjono, dan G.J. Verkerke, "A Framework for Human Spine Imaging Using a Freehand 3D Ultrasound System," *Technol. Heal. Care.*, Vol. 18, No. 1, hal. 1–17, 2010.
- [13] N. Baka, S. Leenstra, dan T. van Walsum, "Random Forest-Based Bone Segmentation in Ultrasound," *Ultrasound Med. Biol.*, Vol. 43, No. 10, hal. 2426–2437, 2017.
- [14] N. Quader, A. Hodgson, dan R. Abugharbieh, *Confidence Weighted*

- Local Phase Features for Robust Bone Surface Segmentation in Ultrasound*, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Cham, Switzerland: Springer, 2014, Vol. 8680, hal. 76–83.
- [15] R. Jia, S.J. Mellon, S. Hansjee, A.P. Monk, D.W. Murray, dan J.A. Noble, "Automatic Bone Segmentation in Ultrasound Images Using Local Phase Features and Dynamic Programming," *IEEE 13th Int. Symp. Biomed. Imaging*, 2016, hal. 1005–1008.
- [16] F. Berton, F. Cheriet, M.C. Miron, dan C. Laporte, "Segmentation of The Spinous Process and Its Acoustic Shadow in Vertebral Ultrasound Images," *Comput. Biol. Med.* Vol. 72, hal. 201–211, 2016.
- [17] L. Lopez-Perez, J. Lemaitre, A. Alfiansyah, dan M.-E. Bellemare, "Bone Surface Reconstruction Using Localized Freehand Ultrasound Imaging," *30th Annual International IEEE EMBS Conference*, 2008, hal. 2964–2967.
- [18] I. Hacıhaliloglu, R. Abugharbieh, A.J. Hodgson, dan R.N. Rohling, "Bone Surface Localization in Ultrasound Using Image Phase-Based Features," *Ultrasound Med. Biol.*, Vol. 35, No. 9, hal. 1475–1487, 2009.
- [19] D. Yang, S. Zhang, Z. Yan, C. Tan, K. Li, dan D. Metaxas, "Automated Anatomical Landmark Detection on Distal Femur Surface Using Convolutional Neural Network," *Proc. - Int. Symp. Biomed. Imaging*, 2015, hal. 17–21.
- [20] H. Ravishankar, S.M. Prabhu, V. Vaidya, dan N. Singhal, "Hybrid Approach for Automatic Segmentation of Fetal Abdomen from Ultrasound Images Using Deep Learning," *Proc. - Int. Symp. Biomed. Imaging*, 2016, hal. 779–782.
- [21] J.C. Nascimento dan G. Carneiro, Multi-Atlas Segmentation Using Manifold Learning with Deep Belief Networks, *Proc. - Int. Symp. Biomed. Imaging*, 2016, hal. 867–871.
- [22] G. Carneiro dan J.C. Nascimento, "Combining Multiple Dynamic Models and Deep Learning Architectures for Tracking the Left Ventricle Endocardium in Ultrasound Data," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 35, No. 11, hal. 2592–2607, 2013.
- [23] Y. Gao, M.A. Maraci, dan J.A. Noble, "Describing Ultrasound Video Content Using Deep Convolutional Neural Networks," *Proc. - Int. Symp. Biomed. Imaging*, 2016, hal. 787–790.
- [24] P.M.B. Torres, J.M. Sanches, P.J.S. Goncalves, dan J.M.M. Martins, "3D Femur Reconstruction Using a Robotized Ultrasound Probe," *Proc. IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics.*, 2012, hal. 884–888.
- [25] S. Ren, K. He, R. Girshick, dan J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *ArXiv Prepr. ArXiv1506.01497.*, Vol. 74, hal. 1–14, 2015.
- [26] (2018) "Colaboratory - Frequently Asked Questions," [Online] <https://research.google.com/colaboratory/faq.html>, tanggal akses: 12-Nov-2018.
- [27] T. Carneiro, R. Victor, M. Da, T. Nepomuceno, G. Bian, dan V.H.C.D.E. Albuquerque, "Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications," *IEEE Access Trends, Perspect. Prospect. Mach. Learn. Appl. to Biomed. Syst. Internet Med. Things.*, Vol. 6, hal. 61677–61685, 2018.
- [28] H. Gao, (2017) "Faster R-CNN Explained - Medium," [Online] <https://medium.com/@smallfishbigsea/faster-r-cnn-explained-864d4fb7e3f8>, tanggal akses: 12-Nov-2018.