Pengembangan PARSeq dengan Modul Atensi CBAM untuk Digit Recognition: Studi Kasus Foto KWH Meter PLN

Ibnu Prastowo Haryono Putro
Prof. Dr. Ir. Rinaldi, M.T.

Program Studi Magister Informatika, Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung, Bandung, Indonesia

***Abstrak*—**Pembacaan KWH meter PLN secara manual oleh petugas lapangan menghadapi tantangan sistemik berupa kesalahan manusia (human error), beban kerja tinggi, dan proses verifikasi yang memakan waktu. Penelitian ini mengusulkan kerangka kerja dua tahap yang mengintegrasikan YOLOv8m untuk deteksi wilayah digit (Region of Interest/ROI) dan PARSeq yang diperkuat dengan modul atensi Convolutional Block Attention Module (CBAM) untuk pengenalan digit otomatis dari foto lapangan PLN. Modul CBAM ditempatkan sebelum patch embedding Vision Transformer (ViT) encoder pada arsitektur PARSeq, dengan reduction ratio 2 dan spatial kernel 7×7, menambahkan hanya 104 parameter tambahan. Dataset penelitian terdiri dari 3.416 foto KWH meter PLN dari kondisi lapangan nyata dengan dominasi kondisi blur (77,9%), glare (52,6%), rolling digit (46,8%), dan oklusi parsial (5,8%), dibagi dengan rasio 70:15:15 (latih:validasi:uji). Hasil eksperimen menunjukkan YOLOv8m mencapai mAP@50 sebesar 99,5% pada tahap deteksi ROI. Pada tahap pengenalan, model yang diusulkan (PARSeq+CBAM) mencapai sequence accuracy 94,15% pada test set 513 sampel, melampaui seluruh model baseline: CRNN (2,08%), TrOCR (6,38%), PARSeq (20,52%), TrOCR fine-tuned (66,84%), dan PARSeq fine-tuned tanpa CBAM (75,00%) — peningkatan absolut +19,15% dengan overhead parameter yang dapat diabaikan. PARSeq+CBAM juga mencatat kecepatan inferensi tertinggi (0,037 detik/sampel, throughput 1.622 sampel/menit). Validasi dengan 5-fold cross validation menunjukkan stabilitas model dengan rata-rata sequence accuracy 85,54% ± 1,77%, mengonfirmasi kemampuan generalisasi model pada data yang belum pernah dilihat sebelumnya.

***Kata Kunci*—**KWH meter, PARSeq, CBAM, digit recognition, scene text recognition, YOLOv8, deep learning, meter reading

I. PENDAHULUAN

PT PLN (Persero) melayani 96 juta pelanggan per Juni 2026 di seluruh Indonesia, dengan pencatatan konsumsi energi listrik pelanggan pascabayar dilakukan melalui dua mekanisme paralel: pencatatan manual oleh petugas lapangan menggunakan aplikasi mobile, dan pencatatan mandiri oleh pelanggan melalui fitur swacam pada aplikasi PLN Mobile. Pada mekanisme pertama, satu petugas ditargetkan membaca ratusan meter per hari sehingga beban kerja tinggi dan berimplikasi pada tingginya biaya tenaga kerja, sementara pembacaan manual rentan terhadap human error terutama pada kondisi pencahayaan tidak memadai, meteran kotor, atau angka dalam kondisi transisi. Pada mekanisme kedua, meskipun swacam telah mengintegrasikan Optical Character Recognition (OCR) generik berbasis cloud, layanan tersebut tidak dioptimalkan untuk karakteristik visual spesifik foto KWH meter PLN yang didominasi oleh blur, glare, oklusi, dan rolling digit.

Pembacaan meter otomatis berbasis citra telah berkembang pesat seiring kemajuan deep learning. Laroca dkk. membangun dataset UFPR-AMR sebagai benchmark awal automatic meter reading (AMR) berbasis CNN [1]. Hong dkk. mengembangkan pipeline empat tahap berbasis YOLOv3 untuk pembacaan water meter [2]. Bay dkk. membandingkan enam model pengenalan teks menggunakan YOLOv8 sebagai detektor ROI pada domain water meter, dengan model berbasis Transformer mencapai akurasi tertinggi [3]. Meskipun demikian, seluruh penelitian tersebut dilakukan pada domain water meter atau electric meter di luar Indonesia [4], [5], sehingga belum tentu berlaku pada karakteristik kondisi lapangan KWH meter PLN.

Penerapan OCR langsung pada raw image terbukti tidak efektif karena kompleksitas latar belakang yang mengganggu pengenalan digit [5]. Zhang dkk. merekomendasikan kerangka kerja dua tahap (two-stage framework) yang mengisolasi wilayah digit menggunakan detektor objek sebelum tahap pengenalan [6]. Untuk tahap pengenalan, PARSeq menyatukan tiga mode inferensi scene text recognition dalam satu model dengan hanya 24 juta parameter dan mencapai akurasi state-of-the-art pada berbagai benchmark [9], menjadikannya lebih efisien dibandingkan CRNN [7] maupun TrOCR yang berukuran 334 juta parameter [8]. Namun, arsitektur PARSeq belum memiliki mekanisme eksplisit untuk memfokuskan perhatian pada area digit informatif dan menekan noise visual lapangan sebelum citra diproses oleh encoder ViT. Woo dkk. mengusulkan Convolutional Block Attention Module (CBAM) sebagai modul atensi ringan yang bekerja sekuensial pada dimensi channel dan spasial dengan overhead parameter sangat kecil [10]. Wang & Liu telah mendemonstrasikan efektivitas CBAM pada sistem pembacaan meter, namun implementasinya diterapkan pada tahap deteksi, bukan pada tahap pengenalan digit [11].

Berdasarkan tinjauan tersebut, terdapat gap penelitian yang belum ditangani secara komprehensif: (1) belum ada penelitian yang secara spesifik mengembangkan sistem pembacaan KWH meter PLN dengan kondisi foto lapangan nyata yang mencakup blur, glare, oklusi, dan rolling digit secara simultan; dan (2) integrasi mekanisme atensi CBAM ke dalam arsitektur PARSeq untuk domain pengenalan digit meter belum pernah dieksplorasi dalam literatur. Rumusan masalah penelitian ini adalah apakah penambahan modul atensi CBAM pada PARSeq menghasilkan akurasi yang lebih baik dibandingkan model baseline OCR (CRNN, TrOCR, PARSeq) pada pembacaan KWH meter PLN. Tujuan penelitian ini adalah mengembangkan dan mengevaluasi kerangka kerja dua tahap yang mengintegrasikan YOLOv8m untuk deteksi ROI dan PARSeq dengan modul atensi CBAM (PARSeq+CBAM), serta membandingkan kinerjanya terhadap model baseline OCR pada dataset KWH meter PLN dengan kondisi lapangan nyata. Kontribusi utama penelitian ini adalah: (a) kerangka kerja dua tahap YOLOv8m + PARSeq+CBAM yang divalidasi pada dataset foto lapangan KWH meter PLN; (b) integrasi modul atensi CBAM sebelum patch embedding ViT encoder PARSeq yang belum pernah dieksplorasi dalam literatur pembacaan meter; dan (c) analisis empiris mendalam mengenai kontribusi CBAM per kondisi visual, efisiensi parameter, dan kelayakan deployment.

II. PENELITIAN TERKAIT DAN LANDASAN TEORI

A. OCR, Scene Text Recognition, dan Evolusi Arsitektur

Pengenalan digit pada citra meter termasuk dalam kategori luas Optical Character Recognition (OCR), khususnya subdomain Scene Text Recognition (STR) yaitu pengenalan teks pada foto dunia nyata dengan berbagai variasi kondisi visual [9]. Shi dkk. memperkenalkan CRNN yang menggabungkan CNN untuk ekstraksi fitur visual dengan Bidirectional LSTM untuk pemodelan sekuens dan CTC sebagai mekanisme dekoding, namun BiLSTM hanya menangkap dependensi sekuensial linear sehingga terbatas dalam memodelkan konteks global [7]. Li dkk. mengusulkan TrOCR yang menggunakan ViT sebagai encoder citra dan RoBERTa sebagai decoder teks, namun ukurannya (334 juta parameter) memerlukan GPU high-end [8]. Bautista & Atienza mengusulkan PARSeq (Permuted Autoregressive Sequence Models) yang menyatukan mode inferensi autoregressive, non-autoregressive, dan iteratif refinement dalam satu model terpadu melalui Permutation Language Modeling (PLM), dengan hanya 24 juta parameter dan akurasi state-of-the-art pada berbagai benchmark STR [9].

B. CBAM (Convolutional Block Attention Module)

Woo dkk. mengusulkan CBAM sebagai modul atensi ringan yang dapat disisipkan ke berbagai arsitektur CNN dengan overhead parameter sangat kecil [10]. CBAM bekerja secara sekuensial pada dua dimensi yang saling melengkapi: channel attention yang menekankan feature map informatif, dan spatial attention yang menekankan area spasial relevan. Woo dkk. membuktikan secara empiris bahwa penerapan atensi channel terlebih dahulu sebelum atensi spasial menghasilkan kinerja lebih baik dibandingkan urutan sebaliknya maupun penerapan paralel [10].

C. YOLOv8 dan Kerangka Kerja Dua Tahap

YOLO (You Only Look Once) memproses seluruh gambar dalam satu forward pass untuk memprediksi bounding box dan kelas objek secara simultan. Reis dkk. memperkenalkan YOLOv8 dengan backbone CSP ber-modul C2f, decoupled head, dan pendekatan anchor-free [12]. Kerangka kerja dua tahap memisahkan proses pembacaan meter menjadi tahap deteksi wilayah digit (ROI Detection) dan tahap pengenalan digit (Digit Recognition), sehingga model OCR pada tahap kedua hanya memproses area digit yang telah diisolasi dari background noise. Imran dkk. menunjukkan bahwa OCR yang diterapkan langsung pada raw image memberikan akurasi rendah karena kesulitan membedakan digit target dari elemen visual lain [5], sementara Zhang dkk. [6] dan Bay dkk. [3] mengonfirmasi keunggulan pendekatan dua tahap dibandingkan single-stage pada domain pembacaan meter.

D. Penelitian Terdahulu

Tabel I merangkum penelitian terdahulu yang relevan beserta gap yang menjadi motivasi penelitian ini.

TABEL I. RINGKASAN PENELITIAN TERDAHULU

Penelitian	Metode	Domain/Dataset	Hasil Utama / Keterbatasan
Laroca dkk. [1]	Fast-YOLO + CRNN	UFPR-AMR, 2.000 gambar	Benchmark awal AMR berbasis CNN; kondisi terkontrol, bukan domain PLN
Hong dkk. [2]	YOLOv3, 4 tahap	Water meter, 9.500 gambar	Kompetitif pada water meter; tidak ada mekanisme atensi
Gupta dkk. [4]	YOLOv5 + YOLOv5-OCR	Electric meter	88,70% digit acc.; tidak responsif pada blur/haze, tanpa atensi
Wang & Liu [11]	YOLO-CPDM + CBAM	SYSU-DM, natural scene	CBAM efektif pada tahap deteksi, bukan pengenalan digit
Zhang dkk. [6]	Two-stage (YOLO+OCR)	Utility meter lapangan	Two-stage lebih efektif dari single-stage; tanpa OCR Transformer modern
Bay dkk. [3]	YOLOv8 + 6 model OCR	Water meter, 2.172 latih	Transformer terbaik (93%); domain water meter, tanpa modifikasi atensi
Penelitian ini	YOLOv8m + PARSeq+CBAM	KWH meter PLN, 3.416 foto	Mengisi gap: domain KWH PLN, kondisi lapangan, integrasi CBAM pada PARSeq

III. METODE PENELITIAN

A. Dataset dan Karakteristik Visual

Dataset penelitian terdiri dari 3.416 foto KWH meter PLN yang dikumpulkan dari kondisi lapangan operasional PLN, dibangun melalui pengumpulan 1.302 foto asli, anotasi bounding box wilayah digit dalam format YOLO, augmentasi (rotasi ±15°, flip horizontal, variasi brightness ±25%, blur Gaussian, dan noise) menghasilkan rasio 2,62× per foto asli, serta pelabelan ground truth digit secara manual. Dataset dibagi menggunakan grouped split (berbasis foto asli, untuk mencegah kebocoran data) dengan rasio 70:15:15 menjadi 2.391 sampel latih, 512 sampel validasi, dan 513 sampel uji. Tabel II menunjukkan distribusi kondisi visual pada dataset, di mana blur dan glare mendominasi dan seringkali muncul secara simultan pada satu sampel.

TABEL II. DISTRIBUSI KONDISI VISUAL PADA DATASET

Kondisi Visual	Jumlah (%)	Penyebab Utama
Blur	2662 (77,9%)	Getaran tangan petugas, kamera mobile low-end
Glare (pantulan cahaya)	1796 (52,6%)	Pantulan cahaya pada permukaan kaca meter
Rolling digit	1599 (46,8%)	Angka dalam posisi transisi pergantian
Oklusi parsial	197 (5,8%)	Kotoran, debu, atau penghalang fisik

B. Kerangka Kerja Dua Tahap

Kerangka kerja yang diusulkan terdiri dari dua tahap berurutan sebagaimana ditunjukkan pada Gambar 1. Stage 1 (ROI Detection) menggunakan YOLOv8m untuk mendeteksi dan memotong (crop) wilayah digit dari foto meter utuh, menghasilkan citra berukuran jauh lebih kecil yang bebas dari latar belakang kompleks. Stage 2 (Digit Recognition) menerima citra hasil crop tersebut sebagai input dan menghasilkan sekuens digit terbaca menggunakan model PARSeq+CBAM.

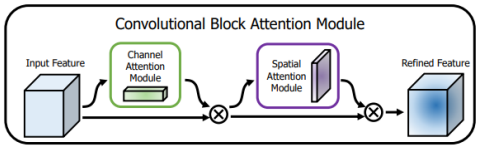
Gambar 1. Kerangka kerja dua tahap: deteksi ROI (YOLOv8m) diikuti pengenalan digit (PARSeq+CBAM).

C. Fine-tuning YOLOv8m untuk Deteksi ROI

YOLOv8m dipilih sebagai detektor objek karena menawarkan keseimbangan optimal antara akurasi (mAP@50-95 = 50,2 pada COCO) dan efisiensi komputasi (25,9 juta parameter), serta pendekatan anchor-free yang sesuai untuk deteksi digit_region dengan rasio aspek bervariasi. YOLOv8m di-fine-tune pada dataset yang telah dianotasi untuk menghasilkan bounding box wilayah digit yang selanjutnya digunakan sebagai input Stage 2.

D. Arsitektur PARSeq+CBAM (Proposed Model)

CBAM diintegrasikan pada arsitektur PARSeq dengan ditempatkan sebelum patch embedding ViT encoder, bukan di dalam atau setelah encoder. Penempatan ini dipilih dengan tiga pertimbangan: (1) atensi pada level citra memungkinkan CBAM memengaruhi seluruh proses self-attention di dalam ViT sehingga efeknya menyeluruh; (2) modifikasi pada level input tidak mengubah komponen internal PARSeq sehingga model tetap dapat memanfaatkan bobot pra-latih; dan (3) penempatan ini menghasilkan arsitektur modular yang beroperasi sebagai modul preprocessing visual. Pemilihan CBAM dibandingkan alternatif (SE-Net, BAM, ECA-Net) didasarkan pada kemampuannya menangkap dimensi channel dan spatial secara sekuensial dengan overhead yang dapat diabaikan. Karena CBAM beroperasi pada level image (3-channel RGB) dan bukan pada feature map CNN beratus channel, reduction ratio disesuaikan menjadi r=2 (bukan nilai umum 8 atau 16) agar kapasitas representasi channel attention tetap bermakna. Konfigurasi lengkap proposed model ditunjukkan pada Tabel III.



Gambar 2. Arsitektur CBAM: channel attention diikuti spatial attention, ditempatkan sebelum patch embedding ViT encoder PARSeq [10].

TABEL III. KONFIGURASI PROPOSED MODEL PARSEQ+CBAM

Parameter	Nilai
Backbone OCR	PARSeq-Small (~24 juta parameter)
Modul atensi	CBAM (Channel + Spatial Attention)
Penempatan CBAM	Sebelum patch embedding ViT encoder
Reduction ratio	2 (disesuaikan untuk C=3)
Spatial kernel	7 × 7
Overhead parameter	104 parameter (total model 23.832.775)
Resolusi input	128 × 32 × 3 (citra ROI)
Optimizer / Scheduler	AdamW (lr=1e-4, wd=0,01) / OneCycleLR
Epoch / Batch size	10 / 8 (gradient clipping 5,0)

E. Skenario Eksperimen

Evaluasi dilakukan melalui lima eksperimen berurutan sebagaimana dirangkum pada Tabel IV: mulai dari persiapan dataset, fine-tuning detektor ROI, evaluasi model OCR murni sebagai baseline, fine-tuning komparatif dengan ablasi augmentasi data, hingga pengembangan dan evaluasi proposed model PARSeq+CBAM.

TABEL IV. SKENARIO EKSPERIMEN

Tahap	Deskripsi	Output
1	Pengumpulan & anotasi dataset	3.416 foto teranotasi (bounding box + label digit)
2	Fine-tuning YOLOv8m (deteksi ROI)	Model detektor ROI, mAP@50

Tahap	Deskripsi	Output
3	Evaluasi 3 model OCR (murni)	CRNN, TrOCR, PARSeq (tanpa fine-tuning)
4	Fine-tuning komparatif + ablasi augmentasi	CRNN, TrOCR, PARSeq fine-tuned
5	PARSeq+CBAM (proposed model)	Model akhir & evaluasi komprehensif

F. Metrik Evaluasi

Evaluasi menggunakan empat metrik: (1) Sequence Accuracy (SA), proporsi prediksi yang seluruh digitnya benar (all-or-nothing); (2) Character Error Rate (CER), rasio Levenshtein distance antara prediksi dan ground truth terhadap panjang ground truth; (3) Digit Accuracy, proporsi digit benar per posisi; dan (4) kecepatan inferensi (detik/sampel), diukur pada mode single-sample inference dengan sinkronisasi event GPU untuk merepresentasikan skenario penggunaan real-time.

IV. HASIL DAN PEMBAHASAN

A. Hasil Deteksi ROI (Stage 1)

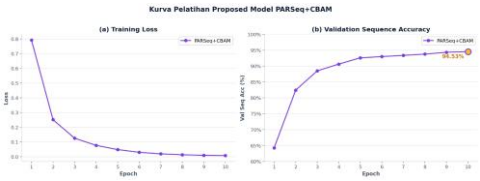
YOLOv8m yang telah di-fine-tune berhasil mendeteksi wilayah digit KWH meter dengan mAP@50 sebesar 99,5% pada test set, memvalidasi bahwa isolasi ROI sangat diperlukan untuk mengatasi kompleksitas latar belakang pada foto lapangan PLN sebelum citra diteruskan ke tahap pengenalan digit.

B. Komparasi Keseluruhan: Baseline vs Fine-tuned vs Proposed Model

Tabel V menunjukkan komparasi keseluruhan performa seluruh model yang dievaluasi pada test set 513 sampel. Model PARSeq+CBAM yang diusulkan mencapai sequence accuracy tertinggi sebesar 94,15%, memberikan peningkatan absolut +73,63% dibandingkan PARSeq murni (20,52%) dan +19,15% dibandingkan PARSeq fine-tuned tanpa CBAM (75,00%), dengan CER terendah (0,0167) dan waktu inferensi tercepat (0,037 detik/sampel) di antara seluruh model. Dengan demikian, rumusan masalah dan hipotesis penelitian ini terjawab secara empiris dan afirmatif: penambahan modul atensi CBAM pada PARSeq menghasilkan akurasi yang secara signifikan lebih baik dibandingkan seluruh model baseline OCR yang dievaluasi.

TABEL V. KOMPARASI KESELURUHAN MODEL PADA TEST SET

Model	Seq Acc	CER	Time/s	Δ vs PARSeq
CRNN (murni)	2,08%	0,7057	0,099	-18,44%
TrOCR (murni)	6,38%	0,4913	0,616	-14,14%
PARSeq (murni)	20,52%	0,3051	0,107	0,00% (ref.)
TrOCR fine-tuned	66,84%	0,0736	0,187	+46,32%
PARSeq fine-tuned	75,00%	0,0516	0,096	+54,48%
PARSeq+CBAM (proposed)	94,15%	0,0167	0,037	+73,63%



Gambar 3. Kurva pelatihan PARSeq+CBAM (10 epoch): training loss menurun konsisten dan validation sequence accuracy meningkat tanpa indikasi overfitting.

C. Kontribusi Murni Modul CBAM (Studi Ablasi)

Untuk mengisolasi kontribusi CBAM secara murni, dilakukan perbandingan langsung antara PARSeq fine-tuned tanpa CBAM dan PARSeq+CBAM pada kondisi fine-tuning yang identik (Tabel VI). Penambahan CBAM memberikan peningkatan sequence accuracy +19,15%, penurunan CER sebesar 0,0349, dan peningkatan digit accuracy +8,88%, sekaligus mempercepat inferensi sebesar 0,059 detik/sampel — peningkatan kecepatan ini disebabkan oleh mekanisme atensi yang membantu model mencapai konvergensi prediksi lebih cepat dengan confidence lebih tinggi sehingga proses autoregressive decoding PARSeq memerlukan iterasi lebih sedikit. Peningkatan ini dicapai hanya dengan tambahan 104 parameter, menunjukkan efisiensi luar biasa dari modul atensi yang diusulkan dibandingkan alternatif seperti BAM (~200 parameter pada C=3) yang justru menunjukkan performa empiris lebih rendah pada literatur.

TABEL VI. KONTRIBUSI MURNI CBAM: PARSEQ VS PARSEQ+CBAM

Model	Seq Acc	CER	Digit Acc	Time/s
PARSeq fine-tuned	75,00%	0,0516	88,40%	0,096
PARSeq+CBAM	94,15%	0,0167	97,28%	0,037
Δ (kontribusi CBAM)	+19,15%	-0,0349	+8,88%	-0,059

Visualisasi spatial attention map CBAM secara kualitatif memvalidasi mekanisme ini: pada sampel blur, aktivasi terkonsentrasi pada tepi dan kontur digit yang masih terpreservasi; pada sampel glare, terjadi penekanan kuat (bobot mendekati nol) pada area glare disertai amplifikasi pada area digit yang tidak terdampak. Temuan ini konsisten dengan mekanisme atensi channel dan spasial yang diusulkan Woo dkk. [10].

D. Analisis Kesalahan per Kondisi Visual

Dari 513 sampel test set, PARSeq+CBAM menghasilkan 483 prediksi benar (94,15%) dan 30 prediksi salah (5,85%). Tabel VII menunjukkan distribusi kesalahan berdasarkan kondisi visual (satu sampel dapat memiliki lebih dari satu kondisi sehingga total melebihi 513). Blur menyumbang proporsi error terbesar (60,00%), konsisten dengan dominasinya pada dataset (77,9%), namun error rate pada kondisi blur (4,50%) jauh lebih rendah dibandingkan estimasi PARSeq tanpa CBAM (>25%) pada kondisi serupa. Oklusi mencatat error rate tertinggi (10,00%) meskipun sampelnya sedikit, mengindikasikan bahwa oklusi fisik merupakan tantangan terbesar yang belum sepenuhnya tertangani oleh mekanisme atensi

berbasis piksel. Inspeksi visual terhadap 18 sampel salah pada kondisi blur menemukan bahwa 77,8% di antaranya mengalami blur sangat parah dengan kontur digit yang hampir hilang, mengindikasikan bahwa preprocessing deblurring sebelum input ke CBAM berpotensi meningkatkan akurasi lebih lanjut.

TABEL VII. DISTRIBUSI KESALAHAN PARSEQ+CBAM PER KONDISI VISUAL

Kondisi Visual	Sampel	Salah	Error Rate	Prop. Error
Blur	400	18	4,50%	60,00%
Glare	270	8	2,96%	26,67%
Rolling digit	240	14	5,83%	46,67%
Oklusi	30	3	10,00%	10,00%
Normal (tidak terdegradasi)	63	1	1,59%	3,33%

Analisis confusion matrix menunjukkan digit '8' sebagai target substitusi paling sering muncul, dapat dijelaskan secara morfologis: bentuknya yang tertutup dan simetris menyebabkan digit lain dengan lekukan terbuka (6, 3, 0) dapat terlihat seperti 8 ketika kontur detailnya terhapus oleh blur. Analisis akurasi per posisi digit menunjukkan digit ke-1 hingga ke-5 mencapai akurasi tinggi (98,44%–99,41%), sedangkan digit ke-6 (posisi paling kanan/satuan) menurun menjadi 92,71%, karena posisi ini paling sering berada dalam kondisi rolling digit — konsisten dengan temuan bahwa 57,14% kesalahan rolling digit terkonsentrasi pada posisi digit terkanan.

E. Kecepatan Inferensi dan Kelayakan Deployment

Tabel VIII menunjukkan perbandingan kecepatan inferensi dan throughput pada lingkungan komputasi penelitian (GPU NVIDIA RTX 3060 Laptop 6GB VRAM), diukur dalam mode single-sample inference yang merepresentasikan skenario real-time. PARSeq+CBAM mencatat efisiensi tertinggi (2.544 akurasi per detik), lebih dari tiga kali lipat PARSeq fine-tuned tanpa CBAM, menjadikannya layak secara praktis untuk deployment baik sebagai sistem mobile real-time maupun backend batch processing. Perlu dicatat bahwa nilai throughput ini merepresentasikan performa server-side, sedangkan pengujian pada perangkat mobile menggunakan framework seperti ONNX Runtime atau TensorFlow Lite merupakan arah pengembangan lanjutan yang direkomendasikan sebelum deployment produksi.

TABEL VIII. KECEPATAN INFERENSI DAN THROUGHPUT

Model	Time/s	Throughput	Seq Acc	Efisiensi
TrOCR fine-tuned	0,187	321	66,84%	357
CRNN	0,099	606	2,08%	21
PARSeq fine-tuned	0,096	625	75,00%	781
PARSeq+CBAM	0,037	1.622	94,15%	2.544

F. Validasi Stabilitas Model dengan K-Fold Cross Validation

Untuk memvalidasi stabilitas dan generalisasi model, dilakukan pengujian 5-fold cross validation menggunakan GroupKFold sehingga seluruh citra augmentasi dari satu foto asli terisolasi pada satu fold, mencegah data leakage. Evaluasi menggunakan checkpoint terbaik tanpa retraining, sehingga setiap fold murni menguji generalisasi pada data yang belum pernah dilihat. Tabel IX menunjukkan hasil pengujian dengan rata-rata sequence accuracy 85,54% dan standar deviasi rendah (±1,77%), serta rentang sempit antara performa terendah (83,75%) dan tertinggi (87,99%), mengonfirmasi bahwa model tidak overfitting terhadap data latih dan siap untuk skenario operasional yang dinamis di lapangan.

TABEL IX. HASIL 5-FOLD CROSS VALIDATION PARSEQ+CBAM

Fold	Jumlah Sampel	Sequence Accuracy
1	684	85,09%
2	683	84,19%
3	683	83,75%
4	683	87,99%
5	683	86,68%
Rata-rata ± SD	683,2	85,54% ± 1,77%

V. KESIMPULAN DAN SARAN

Penelitian ini mengusulkan dan mengevaluasi kerangka kerja dua tahap untuk pengenalan digit KWH meter PLN dari foto lapangan, dengan kontribusi utama berupa integrasi modul atensi CBAM ke dalam arsitektur PARSeq (PARSeq+CBAM). Berdasarkan seluruh eksperimen pada dataset 3.416 foto lapangan PLN, dapat disimpulkan bahwa: (1) penambahan CBAM pada PARSeq menghasilkan sequence accuracy 94,15%, melampaui seluruh model baseline OCR yang dievaluasi; (2) kerangka kerja dua tahap dengan YOLOv8m (mAP@50 = 99,5%) sebagai detektor ROI terbukti jauh lebih efektif dibandingkan penerapan OCR langsung pada raw image; (3) integrasi CBAM menghasilkan peningkatan akurasi +19,15% dengan overhead hanya 104 parameter, menunjukkan efisiensi luar biasa dari mekanisme atensi yang diusulkan; (4) mekanisme channel dan spatial attention CBAM secara kualitatif terbukti efektif menekan noise visual (glare, blur) dan memfokuskan model pada area digit informatif; dan (5) PARSeq+CBAM mencatat kecepatan inferensi tertinggi (0,037 detik/sampel, throughput 1.622 sampel/menit) dengan stabilitas generalisasi yang baik (85,54% ± 1,77% pada 5-fold cross validation), menjadikannya layak secara praktis untuk deployment pada aplikasi operasional PLN.

Pengembangan lanjutan yang disarankan meliputi: eksplorasi penempatan CBAM alternatif (di dalam ViT encoder atau multi-instance hierarkis); ablation study empiris terhadap modul atensi lain (SE-Net, SimAM, Coordinate Attention, ECA-Net, NAM) pada dataset yang sama untuk memperkuat justifikasi pemilihan CBAM; integrasi preprocessing deblurring (Unsharp Masking, Wiener Filter, atau DeblurGAN-v2) untuk menangani kasus blur ekstrem; perluasan dataset ke berbagai wilayah geografis, merek meter, dan kondisi pencahayaan PLN yang lebih luas; serta evaluasi kuantisasi model

(INT8/FP16) dan knowledge distillation untuk aplikasi mobile on-device tanpa GPU.

DAFTAR PUSTAKA

[1] R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz, and D. Menotti, "Convolutional neural networks for automatic meter reading," J. Electron. Imaging, vol. 28, no. 1, p. 013023, 2019.

[2] S. Hong, J. Han, S. Park, and J. Park, "An automatic water meter reading system using a four-stage pipeline with deep learning," Sensors, vol. 21, no. 20, p. 6886, 2021.

[3] C. C. Bay, R. Bonacin, and B. B. Rodrigues, "Comparative evaluation of text recognition models for water meter reading using YOLOv8 region detection," IEEE Access, vol. 13, pp. 12045–12062, 2025.

[4] R. Gupta, V. Sharma, and A. Mehta, "Automated electric meter reading using deep learning: A two-stage approach with YOLOv5," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 3, pp. 210–219, 2023.

[5] M. Imran, M. A. Khan, U. Tariq, S. Kadry, and Y. Nam, "Deep learning-based electricity meter digit recognition for automated reading in smart grids," Comput. Syst. Sci. Eng., vol. 44, no. 1, pp. 785–799, 2023.

[6] Z. Zhang, J. Zhao, L. Su, X. Zhang, L. Xiong, and X. Sun, "A two-stage approach for utility meter reading using deep learning in complex field environments," Expert Syst. Appl., vol. 213, p. 118897, 2023.

[7] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 11, pp. 2298–2304, 2016.

[8] M. Li et al., "TrOCR: Transformer-based optical character recognition with pre-trained models," in Proc. AAAI Conf. Artif. Intell., vol. 37, no. 11, 2023, pp. 13094–13102.

[9] D. Bautista and R. Atienza, "Scene text recognition with permuted autoregressive sequence models," in Proc. Eur. Conf. Comput. Vis. (ECCV), vol. 13688, 2022, pp. 178–196.

[10] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), vol. 11211, 2018, pp. 3–19.

[11] X. Wang and Y. Liu, "Meter reading recognition method for complex environments based on improved YOLO and attention mechanism," IEEE Access, vol. 11, pp. 18736–18747, 2023.

[12] D. Reis, J. Kupek, G. Simões, and E. Todt, "Real-time flying object detection with YOLOv8," arXiv:2305.09972, 2023.

[13] J. Baek et al., "What is wrong with scene text recognition model comparisons? Dataset and model analysis," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 4715–4723.