Jurnal KomtekInfo
https://jkomtekinfo.org/ojs
2024 Vol. 11 No. 3 Hal: 181-186 p-ISSN: 2356-0010, e-ISSN: 2502-8758

Evaluasi Keandalan Peta Saliensi Grad-CAM dan Perilaku Galat Lintas Stratum Luminansi pada Klasifikasi Lubang Jalan

Agung Ramadhanu[1]✉, Maheza Novrayuda[2]

[1]Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK, Jl. Raya Lubuk Begalung, 25221, Padang, Indonesia

[2]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK, Jl. Raya Lubuk Begalung, 25221, Padang, Indonesia

✉ agung_ramadhanu@upiyptk.ac.id

Abstract

Grad-CAM heatmaps are often shown as evidence that a road-damage classifier looks at the damage itself, yet such maps are rarely tested for dependence on the learned parameters. This study examines Grad-CAM for EfficientNet-B0 and MobileNetV3-Large, each trained with three augmentation strategies on 676 unique pothole and normal-road images (five-fold cross-validation, three seeds). Using the model parameter randomization test of Adebayo et al., maps from trained models were compared with maps after randomizing the classification head or all parameters, on 12 test images from fold 0 (seed-42 models). Maps of clean and shadowed images (severity 3) were also compared. Agreement was measured with Spearman rank correlation. For MobileNetV3-Large, full randomization reduced the mean correlation to 0.016–0.045 and head randomization to 0.21–0.36, whereas clean-versus-shadow maps stayed at 0.959–0.968. For EfficientNet-B0, shadow stability was 0.953–0.975 and head randomization gave 0.14–0.21, but full randomization produced all-zero maps, so the correlation was undefined in all 36 comparisons per strategy. Of 239 misclassifications on clean images (18 model configurations), the medium-luminance stratum had the highest error rate for the illum strategy (3.11–3.41%, mostly false positives), against 1.03–1.64% in other strata, and three normal images were misclassified by every configuration. The results indicate that MobileNetV3-Large maps depend on learned weights, but the small sample (12 images, one fold, one seed) and the absence of causal tests limit the conclusions.

Keywords: Explainable AI, Grad-CAM, sanity check, Spearman rank correlation, pothole classification

Abstrak

Peta panas Grad-CAM sering ditampilkan sebagai bukti bahwa pengklasifikasi kerusakan jalan memperhatikan kerusakan itu sendiri, padahal ketergantungan peta pada parameter yang dipelajari model jarang diuji. Penelitian ini memeriksa Grad-CAM pada EfficientNet-B0 dan MobileNetV3-Large, masing-masing dilatih dengan tiga strategi augmentasi pada 676 citra unik lubang jalan dan jalan normal (validasi silang lima lipatan, tiga nilai awal acak atau seed). Dengan uji pengacakan parameter model dari Adebayo dkk., peta model terlatih dibandingkan dengan peta setelah kepala klasifikasi (classification head) atau seluruh parameter diacak, pada 12 citra uji lipatan 0 (model seed 42). Peta citra bersih dan citra berbayang (tingkat keparahan 3) juga dibandingkan. Kesesuaian diukur dengan korelasi peringkat Spearman. Pada MobileNetV3-Large, pengacakan penuh menurunkan rata-rata korelasi menjadi 0,016–0,045 dan pengacakan kepala klasifikasi menjadi 0,21–0,36, sedangkan peta citra bersih terhadap citra berbayang tetap 0,959–0,968. Pada EfficientNet-B0, stabilitas terhadap bayangan 0,953–0,975 dan pengacakan kepala klasifikasi menghasilkan 0,14–0,21, tetapi pengacakan penuh menghasilkan peta bernilai nol seluruhnya sehingga korelasi tidak terdefinisi pada 36 dari 36 perbandingan per strategi. Dari 239 salah klasifikasi pada citra bersih (18 konfigurasi model), stratum luminansi sedang memiliki laju galat tertinggi untuk strategi illum (3,11–3,41%, sebagian besar positif palsu, yaitu false positive), dibanding 1,03–1,64% pada stratum lain, dan tiga citra normal salah diklasifikasi oleh seluruh konfigurasi. Hasil ini menunjukkan bahwa peta MobileNetV3-Large bergantung pada bobot terlatih, tetapi sampel yang kecil (12 citra, satu lipatan, satu seed) dan ketiadaan uji kausal membatasi kesimpulan.

Kata kunci: Explainable AI, Grad-CAM, uji keandalan, korelasi peringkat Spearman, klasifikasi lubang jalan

KomtekInfo is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

1. Pendahuluan

Model deep learning untuk inspeksi jalan semakin sering dipakai untuk membantu keputusan perawatan yang berkaitan dengan keselamatan [1]–[3]. Akurasi tinggi pada data uji belum cukup untuk memutuskan apakah model aman diandalkan, terutama bila keputusannya dapat memicu atau menunda perbaikan jalan. Karena itu, peta saliensi seperti Grad-CAM [4] sering ditampilkan sebagai bukti bahwa model memperhatikan kerusakan yang tepat. Pendekatan serupa, yaitu Grad-CAM++, telah dipakai pada deteksi lubang jalan di kondisi minim cahaya [5].

Peta saliensi tidak otomatis mencerminkan apa yang dipelajari model. Adebayo dkk. [6] menunjukkan bahwa beberapa metode saliensi menghasilkan peta yang hampir sama untuk model terlatih dan model berbobot acak sehingga lebih mirip detektor tepi daripada penjelasan. Model juga dapat mencapai akurasi tinggi melalui jalan pintas, misalnya latar atau artefak yang berkorelasi dengan label [7], [8], dan penjelasan post hoc pada model kotak hitam bermasalah untuk keputusan berisiko tinggi [9]. Evaluasi penjelasan sendiri masih menjadi bidang yang aktif: tersedia perangkat evaluasi sistematis [10], perbaikan uji pengacakan parameter [11], hasil ketidakmungkinan teoretis untuk atribusi fitur [12], dan daftar tantangan terbuka [13]. Pada citra medis, peta saliensi pun tidak selalu sejajar dengan lokalisasi oleh pakar [14].

Pada klasifikasi lubang jalan, penyajian Grad-CAM umumnya bersifat kualitatif, dan, sepengetahuan penulis, uji pengacakan parameter serta analisis galat yang

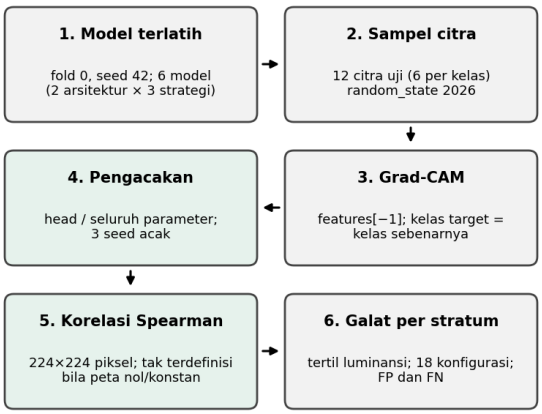
Diterima: 22-09-2024 | Revisi: 24-09-2024 | Diterbitkan: 30-09-2024 | doi: 10.35134/komtekinfo.v11i3.721

dikaitkan dengan pencahayaan belum banyak dilaporkan. Hal ini penting karena bayangan dan cahaya redup adalah penyebab kegagalan yang telah dikenal [15]–[17].

Artikel ini memeriksa Grad-CAM pada dua model konvolusi ringan, EfficientNet-B0 dan MobileNetV3-Large, yang dilatih dengan tiga strategi augmentasi pada 676 citra unik. Pertanyaannya: (RQ1) Apakah peta Grad-CAM bergantung pada parameter terlatih (uji pengacakan parameter model)? (RQ2) Apakah peta stabil ketika citra diberi bayangan sintetis? (RQ3) Pada stratum luminansi mana galat klasifikasi terkonsentrasi, dan citra mana yang salah diklasifikasi secara konsisten? Ketahanan fotometrik pada eksperimen yang sama dilaporkan pada artikel terpisah [25]. Klaim dibatasi pada apa yang dapat ditunjukkan oleh korelasi peringkat; uji ini tidak membuktikan hubungan kausal antara peta dan keputusan model.

2. Metodologi Penelitian

Alur analisis ditunjukkan pada Gambar 1. Seluruh angka dihitung dari prediksi dan peta yang dihasilkan pada protokol eksperimen validasi silang lima lipatan yang diuraikan pada bagian ini; protokol pelatihan yang sama dipakai pada artikel terpisah tentang ketahanan fotometrik [25].



Gambar 1. Alur analisis: pemilihan model dan sampel, perhitungan Grad-CAM, uji pengacakan parameter dengan korelasi Spearman, dan analisis galat per stratum luminansi.

2.1 Data, Model, dan Pelatihan

Data berasal dari *Pothole Detection Dataset* di Kaggle [18]. Dari 681 citra, lima duplikat identik (*hash* SHA-256 atas piksel) dibuang sehingga tersisa 676 citra unik (349 *normal* dan 327 *potholes*); audit ini mencegah kebocoran data antara data latih dan uji [19]. Pencarian citra nyaris duplikat dengan *dHash* tidak menemukan pasangan sehingga setiap citra membentuk satu grup pada *Stratified Group K-Fold* lima lipatan. *Dataset* tidak memiliki metadata lokasi sehingga independensi geografis tidak dapat diverifikasi.

EfficientNet-B0 [20] dan MobileNetV3-Large [21] diinisialisasi dari bobot ImageNet-1K dan dilatih penuh dengan AdamW [22], penjadwalan *cosine annealing*, maksimum 20 *epoch*, dan penghentian dini pada Macro-F1 validasi. Tiga strategi augmentasi dibandingkan: **standard** (pencerminan horizontal), **jitter** (ditambah perubahan kecerahan dan kontras acak), dan **illum** (ditambah gangguan *low-light* atau *shadow* sintetis *severity* 1–3). Pelatihan memakai tiga nilai awal acak (*seed*; 42–44) sehingga ada 90 siklus utama dan prediksi *out-of-fold* untuk setiap citra. Setiap lipatan membagi data menjadi 60% data latih, 20% validasi, dan 20% uji.

2.2 Grad-CAM

Grad-CAM dihitung pada keluaran blok fitur terakhir (*features[−1]*) setiap model [4]. Untuk citra x dan kelas target c, bobot kanal k dan peta kelas adalah

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (1)$$

$$L^c = \mathrm{ReLU}\left(\sum_k \alpha_k^c A^k\right) \quad (2)$$

dengan A^k peta aktivasi kanal k, y^c skor kelas, dan Z jumlah posisi spasial. Peta berukuran 7 × 7 diubah ke 224 × 224 dengan interpolasi bilinear dan dibagi nilai maksimumnya. Kelas target adalah kelas sebenarnya dari citra. Peta dengan nilai maksimum $\leq 10^{-12}$ dicatat sebagai peta nol (*zero map*).

2.3 Sampel dan Model yang Dianalisis

Analisis Grad-CAM memakai model lipatan 0 dengan *seed* 42 (enam model: 2 arsitektur × 3 strategi) dan 12 citra uji lipatan 0, yaitu enam per kelas, yang dipilih secara acak dengan nilai awal pembangkit acak 2026; daftar pengenal citra dicatat untuk menjamin ketertiruan eksperimen. Sampel ini kecil dan hanya mewakili satu lipatan serta satu *seed*.

2.4 Uji Pengacakan Parameter Model

Mengikuti Adebayo dkk. [6], peta model terlatih dibandingkan dengan peta model yang parameternya diacak ulang. Pengujian menerapkan dua tahap: (1) pengacakan kepala klasifikasi (*head randomized*), yaitu inisialisasi ulang modul pengklasifikasi, dan (2) pengacakan seluruh parameter (*all parameters randomized*), yaitu inisialisasi ulang seluruh modul dengan fungsi *reset_parameters*. Setiap tahap diulang dengan tiga *seed* acak (2026–2028) sehingga ada 12 citra × 3 = 36 perbandingan per model. Keutuhan bobot model terlatih setelah pengacakan diverifikasi dengan nilai *hash* SHA-256 pada kamus parameter (*state dict*). Perlu dicatat bahwa ini hanya dua tahap, bukan pengacakan bertahap per lapisan dari atas ke bawah seperti pada makalah asli, dan tidak ada pembanding detektor tepi. Sebagai uji stabilitas, peta citra bersih dibandingkan dengan peta citra yang sama setelah bayangan sintetis *severity* 3 (satu realisasi per citra, 12 perbandingan per model).

2.5 Korelasi Spearman

Kesesuaian dua peta diukur dengan korelasi peringkat Spearman pada seluruh 224 × 224 piksel (n = 50.176), dihitung dengan SciPy [23]:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad (3)$$

dengan d_i selisih peringkat piksel ke-i pada kedua peta. Bila salah satu peta konstan atau tidak berhingga, ρ dinyatakan tidak terdefinisi dan dikeluarkan dari rata-rata;

jumlah perbandingan valid selalu dilaporkan. Tidak ada ambang lulus yang ditetapkan sebelumnya sehingga nilai ρ mendekati nol ditafsirkan secara deskriptif sebagai peta yang berbeda. Rata-rata dan median dilaporkan tanpa selang kepercayaan karena sampelnya kecil [24].

2.6 Analisis Galat berdasarkan Stratum Luminansi

Luminansi rata-rata setiap citra asli dihitung dari luma ITU-R BT.709, yaitu $Y = 0{,}2126\,R + 0{,}7152\,G + 0{,}0722\,B$ pada nilai piksel ternormalisasi [0, 1] setelah citra diperkecil menjadi 128 × 128 piksel. Stratum rendah, sedang, dan tinggi ditentukan dari tertil (*tertile*) luminansi data latih pada tiap lipatan (batas bawah 0,432–0,437 dan batas atas 0,512–0,517), lalu diterapkan pada citra uji lipatan yang sama. Pada prediksi citra bersih dari 18 konfigurasi (2 arsitektur × 3 strategi × 3 *seed*), dihitung laju galat per stratum, *false positive* (*normal* diprediksi lubang), dan *false negative* (lubang diprediksi *normal*). Untuk setiap citra dihitung pula jumlah konfigurasi (maksimum 18) yang salah mengklasifikasikannya. Tabel 2 memuat strategi *illum* (rata-rata tiga *seed*); strategi lain dilaporkan dalam teks.

3. Hasil dan Pembahasan

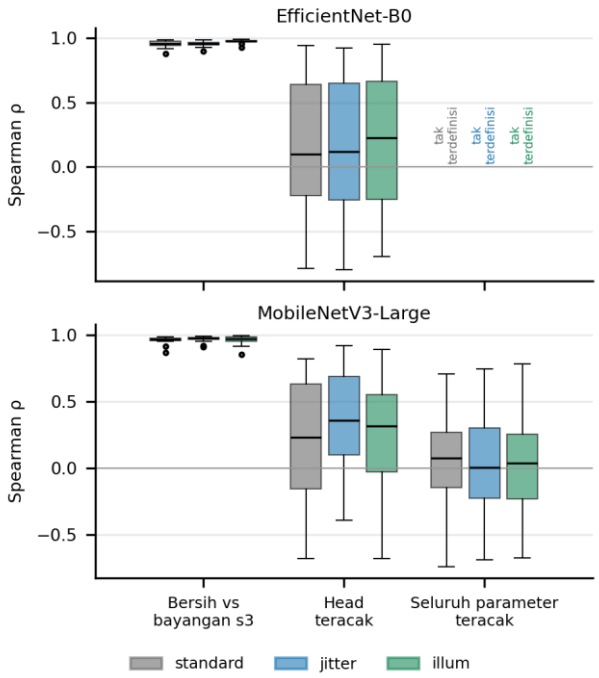
3.1 Uji Pengacakan Parameter

Tabel 1 merangkum korelasi Spearman. Pada MobileNetV3-Large, pengacakan seluruh parameter menurunkan rata-rata ρ menjadi 0,016–0,045 pada ketiga strategi (36 dari 36 perbandingan valid), artinya peta model teracak hampir tidak berkorelasi peringkat dengan peta model terlatih. Pengacakan kepala klasifikasi saja menurunkan ρ menjadi 0,21–0,36; peta berubah banyak tetapi tidak sepenuhnya, wajar karena jaringan tulang punggung (*backbone*) masih terlatih. Hasil ini konsisten dengan peta MobileNetV3-Large yang bergantung pada parameter terlatih sehingga bukan sekadar fungsi dari *input* seperti detektor tepi. Pada satu perbandingan *head randomized* untuk strategi *standard*, peta konstan sehingga hanya 35 dari 36 perbandingan yang valid. Sebaran per perbandingan lebar (Gambar 2): pada pengacakan penuh, rentang antarkuartil kira-kira −0,2 hingga +0,3 untuk ketiga strategi sehingga rata-rata mendekati nol tidak boleh dibaca terlepas dari sebarannya.

Pada EfficientNet-B0, pengacakan kepala klasifikasi menurunkan ρ menjadi 0,14–0,21, mirip dengan MobileNetV3-Large. Namun pada pengacakan seluruh parameter, ρ tidak terdefinisi pada 36 dari 36 perbandingan untuk ketiga strategi. Diagnostik menunjukkan bahwa pada seluruh perbandingan itu peta model teracak bernilai nol sesudah ReLU, sedangkan peta model terlatih tidak pernah nol. Dengan demikian, uji ini tidak menghasilkan angka untuk EfficientNet-B0, dan model tidak dapat dinyatakan lulus atau gagal berdasarkan ρ. Peta yang menghilang sama sekali tidak sesuai dengan perilaku detektor tepi, tetapi kesimpulan itu lebih lemah daripada korelasi mendekati nol. Penjelasan yang paling sederhana adalah bahwa setelah inisialisasi ulang, jumlah berbobot gradien dan aktivasi pada blok terakhir bernilai negatif di semua posisi sehingga ReLU meniadakannya; penyebab ini belum diuji. Blok fitur terakhir EfficientNet-B0 memakai aktivasi SiLU [20], yang mengizinkan nilai aktivasi negatif kecil; hal ini mungkin berkaitan dengan peta nol, tetapi belum diuji. Pengaruh pemrosesan peta terhadap hasil uji pengacakan parameter telah dibahas pada literatur [11]. Peta tanpa ReLU atau metrik kemiripan yang dapat menerima peta bertanda perlu dicoba pada penelitian lanjutan.

Tabel 1. Rata-rata korelasi Spearman antarpeta Grad-CAM (perbandingan valid/total); tanda – berarti tidak terdefinisi. Model lipatan 0, *seed* 42. EffNet-B0 = EfficientNet-B0; MNv3-L = MobileNetV3-Large

Model	Strategi	Bersih vs. bayangan	Kepala diacak	Semua diacak
EffNet-B0	*standard*	0,953 (12/12)	0,146 (36/36)	– (0/36)
EffNet-B0	*jitter*	0,954 (12/12)	0,138 (36/36)	– (0/36)
EffNet-B0	*illum*	0,975 (12/12)	0,206 (36/36)	– (0/36)
MNv3-L	*standard*	0,959 (12/12)	0,206 (35/36)	0,045 (36/36)
MNv3-L	*jitter*	0,968 (12/12)	0,360 (36/36)	0,028 (36/36)
MNv3-L	*illum*	0,960 (12/12)	0,257 (36/36)	0,016 (36/36)



Gambar 2. Sebaran korelasi Spearman per perbandingan dan strategi. Kotak tidak digambar bila seluruh nilai tidak terdefinisi (EfficientNet-B0, seluruh parameter diacak).

3.2 Stabilitas terhadap Bayangan

Korelasi antara peta citra bersih dan peta citra berbayang *severity* 3 adalah 0,959–0,968 untuk MobileNetV3-Large dan 0,953–0,975 untuk EfficientNet-B0, dengan 12 dari 12 perbandingan valid pada setiap model. Oleh karena itu, peta sangat stabil terhadap bayangan sintetis yang menutupi sebagian citra. Stabilitas saja tidak membuktikan apa-apa, karena peta yang tidak peka terhadap model pun akan stabil. Nilainya baru informatif

karena peta yang sama juga peka terhadap parameter terlatih (Subbab 3.1). Gabungan kedua hasil menunjukkan bahwa peta bergantung pada *input* dan pada parameter terlatih, setidaknya pada MobileNetV3-Large.

3.3 Pemeriksaan Visual

Gambar 3 menampilkan dua citra dari sampel untuk MobileNetV3-Large (*illum*), Gambar 4 untuk EfficientNet-B0 (*illum*), dan Gambar 5 untuk MobileNetV3-Large (*standard*). Pada citra lubang (probabilitas lubang 0,999 dan 1,000), peta terkonsentrasi pada area lubang berair dan tumpukan kerikil, dan tetap di sana setelah bayangan *severity* 3. Pada citra kelas *normal*, peta berpusat pada lajur dan kendaraan di tengah jalan raya, bukan pada tekstur aspal. Hal ini wajar karena tidak ada kerusakan untuk dilokalisasi, tetapi berarti peta kelas *normal* menggambarkan konteks adegan, dan citra kelas *normal* pada *dataset* ini tidak semuanya berupa permukaan aspal dari jarak dekat. Pola ini selaras dengan kemungkinan pemanfaatan petunjuk kontekstual (jalan pintas) seperti yang dibahas pada literatur [7], [8], tetapi dugaan itu tidak diuji di sini. Pada citra kelas *normal*, wilayah yang disorot berbeda antarmodel dan antarstrategi: berpusat pada lajur tengah pada MobileNetV3-Large (*illum*), meluas ke lajur kiri pada EfficientNet-B0 (*illum*), dan bergeser ke sisi kanan jalan pada MobileNetV3-Large (*standard*). Pada citra lubang, ketiga peta konsisten pada area yang sama. Dengan demikian, peta kelas *normal* kurang dapat dijadikan bukti penalaran model. Peta berasal dari blok fitur 7 × 7 sehingga beresolusi kasar, dan dua contoh tidak cukup untuk kesimpulan umum.



Gambar 3. Grad-CAM MobileNetV3-Large (*illum*) pada citra bersih dan citra berbayang *severity* 3; baris atas kelas *normal*, baris bawah kelas *potholes*. Kolom dari kiri ke kanan pada Gambar 3–5: citra bersih, Grad-CAM citra bersih, citra berbayang *severity* 3, dan Grad-CAM citra berbayang.



Gambar 4. Grad-CAM EfficientNet-B0 (*illum*) pada citra dan konfigurasi yang sama dengan Gambar 3.



Gambar 5. Grad-CAM MobileNetV3-Large (*standard*) pada citra yang sama.

Perbedaan antarstrategi pada Tabel 1 tidak dapat ditafsirkan dengan sampel 12 citra. Rata-rata ρ untuk pengacakan kepala klasifikasi pada MobileNetV3-Large berkisar 0,21 (*standard*), 0,26 (*illum*), dan 0,36 (*jitter*), dan stabilitas terhadap bayangan pada EfficientNet-B0 lebih tinggi pada *illum* (0,975) daripada pada *standard* dan *jitter* (0,953 dan 0,954). Selisih ini tidak disertai selang kepercayaan dan tidak boleh dibaca sebagai efek strategi augmentasi.

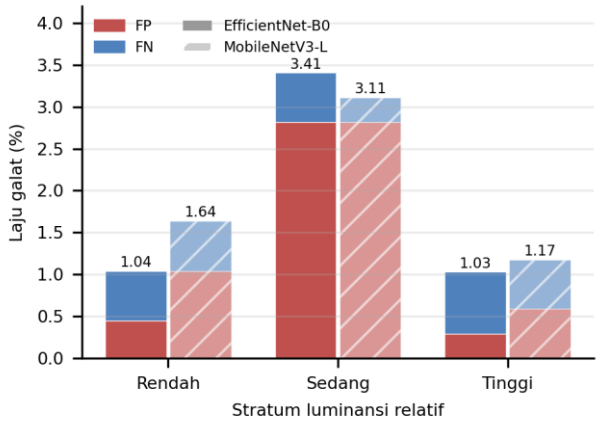
3.4 Analisis Galat pada Citra Bersih

Pada citra bersih terdapat 239 prediksi salah dari 12.168 prediksi (18 konfigurasi × 676 citra), terdiri atas 174 *false positive* dan 65 *false negative*, yang melibatkan 38 citra unik. Pada strategi *illum* (Tabel 2), stratum luminansi sedang, bukan rendah, memiliki laju galat tertinggi: 3,41 ± 0,26% pada EfficientNet-B0 dan 3,11 ± 0,89% pada MobileNetV3-Large, dibanding 1,03–1,64% pada stratum rendah dan tinggi. Pola yang sama muncul pada keenam konfigurasi arsitektur–strategi (laju galat stratum sedang 2,81–3,70%). Sebagian besar galat stratum sedang adalah *false positive*: rata-rata 6,3 dari 7,7 galat per *seed* pada EfficientNet-B0 dan 6,3 dari 7,0 pada MobileNetV3-Large. Secara keseluruhan, 133 dari 239 prediksi salah (56%) berada pada stratum sedang, dan 114 di antaranya adalah citra *normal* yang diprediksi sebagai lubang.

Tabel 2. Galat klasifikasi citra bersih per stratum luminansi (strategi *illum*; rata-rata tiga *seed*)

Model	Stratum	Citra	FP	FN	Laju galat (%)
EffNet-B0	Rendah	224	1,0	1,3	1,04 ± 0,52
EffNet-B0	Sedang	225	6,3	1,3	3,41 ± 0,26
EffNet-B0	Tinggi	227	0,7	1,7	1,03 ± 0,67
MNv3-L	Rendah	224	2,3	1,3	1,64 ± 0,26
MNv3-L	Sedang	225	6,3	0,7	3,11 ± 0,89
MNv3-L	Tinggi	227	1,3	1,3	1,17 ± 0,25

FP: *false positive* (*normal* diprediksi lubang); FN: *false negative*; keduanya rata-rata per *seed*. Laju galat: rata-rata ± simpangan baku tiga *seed*.



Gambar 6. Laju galat per stratum luminansi relatif pada strategi *illum*, dipisah menjadi *false positive* dan *false negative*.

Temuan ini berbeda dari dugaan bahwa citra paling gelap paling sulit, tetapi tidak bertentangan dengan kerentanan terhadap penggelapan sintetis ekstrem pada artikel terpisah [25]: stratum di sini dibentuk dari citra asli tanpa gangguan, sedangkan penurunan terbesar pada artikel tersebut muncul setelah citra digelapkan secara sintetis (*low-light severity* 5). Stratum diturunkan dari luminansi rata-rata citra asli, bukan dari keberadaan bayangan sehingga data ini tidak menjelaskan penyebabnya. Dugaan yang dapat diuji adalah bahwa citra dengan luminansi menengah memuat pencahayaan campuran atau pantulan pada aspal yang menyerupai tepi lubang. Dugaan ini belum diperiksa karena *dataset* tidak memiliki anotasi bayangan dan citra belum ditinjau secara manual.

Tiga citra *normal* (ID 0824261a, ff13bce5, dan 24273036) salah diklasifikasi pada seluruh 18 konfigurasi (Tabel 3); ketiganya berada pada stratum sedang dengan luminansi 0,445–0,500. Kesalahan yang konsisten lintas arsitektur, strategi, dan *seed* lebih mengarah pada citra yang sulit atau berpotensi berlabel keliru daripada kelemahan satu model. Pemeriksaan visual oleh ahli jalan diperlukan sebelum menyimpulkan.

Tabel 3. Delapan citra yang paling konsisten salah diklasifikasi pada citra bersih

ID citra	Kelas	Salah	Stratum	Luminansi
0824261a	*normal*	18/18	Sedang	0,500
ff13bce5	*normal*	18/18	Sedang	0,463
24273036	*normal*	18/18	Sedang	0,445
18c903fd	*normal*	16/18	Sedang	0,437
2be8f06a	*normal*	15/18	Sedang	0,448
42a0df58	*potholes*	14/18	Tinggi	0,538
c292e95e	*normal*	14/18	Rendah	0,397
15cb81cd	*normal*	13/18	Sedang	0,471

ID: 8 karakter awal *hash* piksel; Salah: jumlah konfigurasi (dari 18) yang salah mengklasifikasikan citra.

3.5 Implikasi

Bagi sistem inspeksi jalan, temuan ini merekomendasikan tiga langkah praktis: menjalankan uji pengacakan parameter sebelum peta saliensi dipakai sebagai bukti transparansi, melaporkan jumlah perbandingan yang valid dan bukan hanya rata-rata ρ, serta meninjau citra yang salah secara konsisten sebelum menyalahkan kondisi pencahayaan.

3.6 Keterbatasan Penelitian

Penelitian ini memiliki enam keterbatasan. (1) Ukuran sampel analisis Grad-CAM terbatas pada 12 citra dari satu lipatan dan satu *seed* model sehingga rata-rata ρ dilaporkan tanpa selang kepercayaan. (2) Pengujian pengacakan baru mencakup dua tahap, belum berupa pengacakan berjenjang per lapisan, dan belum ada pembanding detektor tepi maupun ambang lulus yang ditetapkan sebelumnya. (3) EfficientNet-B0 tidak menghasilkan ρ pada pengacakan penuh. (4) Korelasi peringkat tidak membuktikan hubungan kausal, dan Grad-CAM pada blok fitur terakhir beresolusi kasar (7 × 7). (5) Stratum luminansi bukan label bayangan, dan penyebab kegagalan belum diperiksa secara manual. (6) *Dataset* tidak memiliki metadata koordinat GPS ruas jalan sehingga independensi geografis antarlokasi belum dapat diverifikasi secara fisik, dan validasi eksternal belum dilakukan..

4. Kesimpulan

Pada MobileNetV3-Large, korelasi Spearman antara peta Grad-CAM model terlatih dan model teracak penuh mendekati nol (0,016–0,045), dan peta tetap stabil terhadap bayangan sintetis (0,959–0,968); hasil ini mengindikasikan bahwa peta bergantung pada parameter terlatih. Pada EfficientNet-B0, peta stabil terhadap bayangan (0,953–0,975) dan peka terhadap pengacakan kepala klasifikasi (0,14–0,21), tetapi pengacakan penuh menghasilkan peta nol sehingga korelasi tidak terdefinisi dan uji tidak dapat dinilai. Galat klasifikasi terkonsentrasi pada stratum luminansi sedang (3,11–3,41% untuk strategi *illum*), terutama *false positive*, dan tiga citra *normal* salah pada seluruh konfigurasi. Karena sampelnya hanya 12 citra dan tidak ada uji kausal, hasil ini bersifat awal. Penelitian lanjutan perlu menambah jumlah citra dan *seed*, memakai pengacakan bertahap per lapisan dengan pembanding detektor tepi, menguji varian tanpa ReLU untuk EfficientNet-B0, dan meninjau citra yang salah secara manual dengan metadata ruas jalan.

Daftar Rujukan

[1] Y. Safyari, M. Mahdianpari, and H. Shiri, "A review of vision-based pothole detection methods using computer vision and machine learning," Sensors, vol. 24, no. 17, art. 5652, 2024, doi: 10.3390/s24175652.

[2] R. Kothai, N. Prabakaran, Y. V. Srinivasa Murthy, L. R. Cenkeramaddi, and V. Kakani, "Pavement distress detection, classification, and analysis using machine learning algorithms: A survey," IEEE Access, vol. 12, pp. 126943–126960, 2024, doi: 10.1109/ACCESS.2024.3455093.

[3] V. Plevris and G. Papazafeiropoulos, "AI in structural health monitoring for infrastructure maintenance and safety," Infrastructures, vol. 9, no. 12, art. 225, 2024, doi: 10.3390/infrastructures9120225.

[4] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Int. J. Comput. Vis., vol. 128, no. 2, pp. 336–359, 2020, doi: 10.1007/s11263-019-01228-7.

[5] Y. Zanevych, V. Yovbak, O. Basystiuk, N. Shakhovska, S. Fedushko, and S. Argyroudis, "Evaluation of pothole detection performance using deep learning models under low-light

conditions," Sustainability, vol. 16, no. 24, art. 10964, 2024, doi: 10.3390/su162410964.

[6] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 31, 2018, pp. 9505–9515.

[7] S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek, and K.-R. Müller, "Unmasking Clever Hans predictors and assessing what machines really learn," Nat. Commun., vol. 10, art. 1096, 2019, doi: 10.1038/s41467-019-08987-4.

[8] R. Geirhos et al., "Shortcut learning in deep neural networks," Nat. Mach. Intell., vol. 2, no. 11, pp. 665–673, 2020, doi: 10.1038/s42256-020-00257-z.

[9] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat. Mach. Intell., vol. 1, no. 5, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[10] A. Hedström et al., "Quantus: An explainable AI toolkit for responsible evaluation of neural network explanations and beyond," J. Mach. Learn. Res., vol. 24, no. 34, pp. 1–11, 2023. [Online]. Available: https://jmlr.org/papers/v24/22-0142.html

[11] A. Hedström, L. Weber, S. Lapuschkin, and M. M.-C. Höhne, "Sanity checks revisited: An exploration to repair the model parameter randomisation test," in NeurIPS 2023 Workshop XAI in Action: Past, Present, and Future Applications, 2023. [Online]. Available: https://openreview.net/forum?id=vVpefYmnsG

[12] B. Bilodeau, N. Jaques, P. W. Koh, and B. Kim, "Impossibility theorems for feature attribution," Proc. Natl. Acad. Sci. U.S.A., vol. 121, no. 2, art. e2304406120, 2024, doi: 10.1073/pnas.2304406120.

[13] L. Longo et al., "Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions," Inf. Fusion, vol. 106, art. 102301, 2024, doi: 10.1016/j.inffus.2024.102301.

[14] A. Saporta et al., "Benchmarking saliency methods for chest X-ray interpretation," Nat. Mach. Intell., vol. 4, no. 10, pp. 867–878, 2022, doi: 10.1038/s42256-022-00536-x.

[15] L. Fan, S. Li, Y. Li, B. Li, D. Cao, and F.-Y. Wang, "Pavement cracks coupled with shadows: A new shadow-crack dataset and a shadow-removal-oriented crack detection approach," IEEE/CAA J. Autom. Sin., vol. 10, no. 7, pp. 1593–1607, 2023, doi: 10.1109/JAS.2023.123447.

[16] X. Xu, T. Wu, W. Ge, L. Dai, J. Gong, and D. Chen, "Enhanced pothole detection in low-light conditions using DCE-Pothole and YOLOv10 fusion," Results Eng., vol. 28, art. 107322, 2025, doi: 10.1016/j.rineng.2025.107322.

[17] J. Frnda, S. Bandyopadhyay, M. Pavlicko, M. Durica, M. Savrasovs, and S. Banerjee, "Analysis of pothole detection accuracy of selected object detection models under adverse conditions," Transp. Telecommun., vol. 25, no. 2, pp. 209–217, 2024, doi: 10.2478/ttj-2024-0016.

[18] A. Kumar, "Pothole Detection Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/atulyakumar98/pothole-detection-dataset (accessed Oct. 13, 2026).

[19] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," Patterns, vol. 4, no. 9, art. 100804, 2023, doi: 10.1016/j.patter.2023.100804.

[20] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), PMLR 97, 2019, pp. 6105–6114.

[21] A. Howard et al., "Searching for MobileNetV3," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[22] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2019.

[23] P. Virtanen et al., "SciPy 1.0: Fundamental algorithms for scientific computing in Python," Nat. Methods, vol. 17, no. 3, pp. 261–272, 2020, doi: 10.1038/s41592-019-0686-2.

[24] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," Sci. Rep., vol. 14, art. 6086, 2024, doi: 10.1038/s41598-024-56706-x.

[25] M. Novrayuda, "Evaluasi ketahanan fotometrik model jaringan konvolusi ringan melalui augmentasi iluminasi sintetis pada klasifikasi lubang jalan," naskah belum dipublikasikan, 2026..