



Systematic Literature Review : Tren Perkembangan Model dan Algoritma Analisis Video Kerumunan-Padat

Doni Fristiyanto ¹, Abu Khalid Rivai ²

^{1,2} Pascasarjana, Magister Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Banten
Email: ¹ firstiyanto@gmail.com, ² dosen01591@unpam.ac.id

ABSTRACT

Dense-Crowd video analysis is a branch of computer vision that has various important applications in public safety, emergency management, urban planning, pedestrian traffic engineering, and crowd management at large events, such as religious activities, music concerts, and sports matches. This study presents a Systematic Literature Review (SLR) of 30 scientific publications published between 2010 and 2025. The main objective of this review is to identify the latest research trends, classification of algorithms used, application domains, and the main challenges still faced in crowd video analysis. The results of this SLR show that deep learning-based approaches, such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Transformer, still dominate various applications, especially in anomaly detection which aims to recognize suspicious behavior in dense crowds. This technology has significant potential for preventing dangerous events such as riots, mass panic, or accidents. In addition, trends in the integration of new technologies are also found, such as the use of hybrid algorithms that combine several approaches, federated learning for distributed model training, and the use of multimodal data and drones to improve monitoring effectiveness. However, many challenges remain, such as limited representative datasets, decreased accuracy under extreme conditions, computational limitations for real-time applications, and issues of privacy and model interpretability. Therefore, the results of this SLR are expected to make a strategic contribution to the development of more sophisticated, adaptive, and relevant crowd analytics systems.

Keywords: video analysis; dense crowd; deep learning; tracking; systematic literature review

ABSTRAK

Analisis video kerumunan-padat merupakan salah satu cabang dari computer vision yang memiliki berbagai aplikasi penting, seperti dalam bidang keselamatan publik, manajemen keadaan darurat, perencanaan tata kota, rekayasa lalu lintas pejalan kaki, serta pengelolaan kerumunan pada acara-acara besar, seperti kegiatan keagamaan, konser musik, hingga pertandingan olahraga. Penelitian ini menyajikan sebuah *Systematic Literature Review* (SLR) terhadap 30 publikasi ilmiah yang diterbitkan antara tahun 2010 hingga 2025. Tujuan utama dari kajian ini adalah untuk mengidentifikasi tren penelitian terkini, klasifikasi algoritma yang digunakan, domain penerapan, serta tantangan utama yang masih dihadapi dalam analisis video kerumunan. Hasil dari SLR ini menunjukkan bahwa pendekatan berbasis *deep learning*, seperti CNN, LSTM, dan Transformer, masih mendominasi berbagai aplikasi, khususnya dalam deteksi anomali yang bertujuan untuk mengenali perilaku mencurigakan di tengah kerumunan padat. Teknologi ini dinilai potensial dalam mencegah kejadian berbahaya seperti kerusuhan, kepanikan massal, atau kecelakaan. Selain itu, ditemukan pula tren integrasi teknologi baru seperti penggunaan algoritma hibrida yang menggabungkan beberapa pendekatan, federated learning untuk pelatihan model secara terdistribusi, serta pemanfaatan data multimodal dan drone untuk meningkatkan efektivitas pemantauan. Meskipun demikian, masih banyak tantangan yang dihadapi, seperti keterbatasan dataset yang representatif, penurunan akurasi dalam kondisi ekstrem, keterbatasan komputasi untuk aplikasi waktu nyata, serta isu privasi dan interpretabilitas model. Oleh karena itu, hasil SLR ini diharapkan dapat memberikan kontribusi strategis bagi pengembangan sistem analitik kerumunan yang lebih canggih, adaptif, dan relevan.

Kata kunci: analisis video; kerumunan-padat; *deep learning*; pelacakan; *systematic literature review*

1. PENDAHULUAN

Bidang keselamatan publik, manajemen darurat, perencanaan kota, rekayasa lalu lintas pejalan kaki, serta penyelenggaraan acara yang menarik kerumunan besar (seperti acara keagamaan, olahraga, konser musik, dan sejenisnya) sering kali memerlukan analisis kerumunan untuk memastikan pengendalian yang efektif. Kerumunan yang padat biasanya disebabkan oleh pertumbuhan populasi serta peningkatan aktivitas publik di ruang terbuka maupun tertutup. Melalui analisis video, pergerakan individu dapat terdeteksi, kejadian anomali dapat diidentifikasi, individu dapat dilacak, dan risiko dapat diprediksi secara real-time saat kerumunan terjadi. Namun, tantangan teknis yang dihadapi cukup kompleks, mengingat bahwa visualisasi dapat melibatkan tumpang tindih objek, variasi tingkat kepadatan, dan pergeseran perspektif kamera.

Dalam dekade terakhir, pendekatan terhadap analisis kerumunan telah mengalami transformasi signifikan, beralih dari metode tradisional yang hanya mengandalkan fitur visual sederhana, seperti tekstur dan aliran optik, menuju penggunaan algoritma berbasis pembelajaran mendalam yang lebih canggih serta model simulasi yang mampu merepresentasikan perilaku kompleks antar individu. Metode pembelajaran mendalam, termasuk *Convolutional Neural Networks* (CNN), *Recurrent Neural Networks* (RNN), dan Transformers, kini menjadi dominan dalam penelitian terbaru yang berkaitan dengan penghitungan kerumunan, deteksi anomali, prediksi perilaku, dan pelacakan. Untuk memahami perkembangan tren serta inovasi dalam model dan algoritma analisis video kerumunan padat, dapat dilakukan *Systematic Literature Review* (SLR) pada jurnal-jurnal ilmiah yang memuat model dan algoritma analisis video kerumunan padat, yang terindeks dalam basis data publikasi ilmiah berkualitas tinggi dan memiliki reputasi internasional, berkat proses seleksi yang ketat dan standar yang diterapkan.

Penelitian ini menyajikan *Systematic Literature Review* (SLR) mengenai model dan algoritma yang digunakan dalam analisis video kerumunan-padat, berdasarkan dari 30 publikasi ilmiah. Fokus utama dari penelitian ini adalah:

1. mengklasifikasi metode analisis berdasarkan pendekatan yang digunakan (klasik, *deep learning*, simulasi),
2. mengevaluasi kinerja dan kompleksitas dari pendekatan yang diterapkan,
3. mengidentifikasi tren dan inovasi terbaru, serta

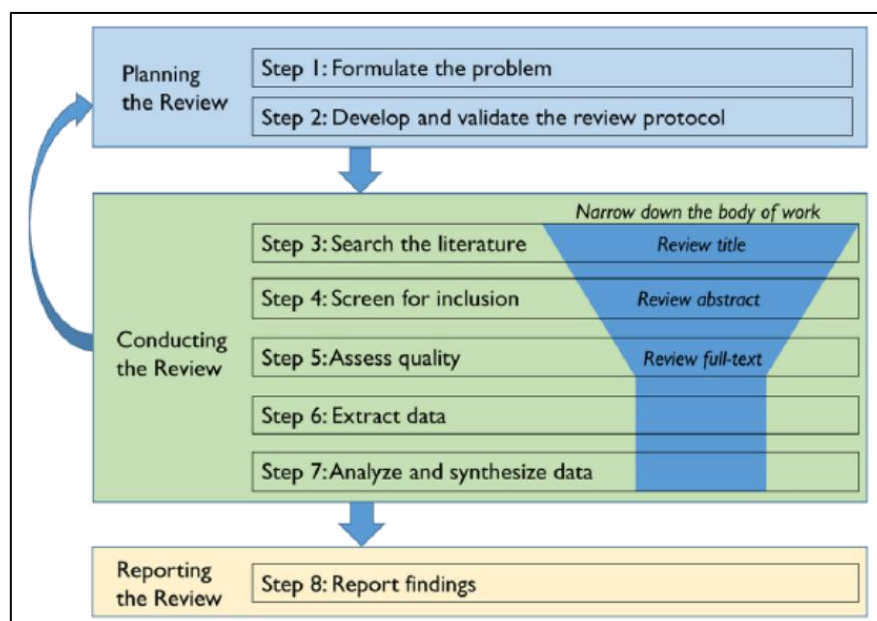
4. menemukan kesenjangan penelitian yang belum banyak dieksplorasi oleh komunitas ilmiah.

Dengan demikian, hasil penelitian ini diharapkan dapat menjadi referensi strategis bagi para peneliti dan praktisi dalam merancang sistem analisis kerumunan yang lebih akurat, adaptif, dan efisien di masa mendatang.

2. METODE

Penelitian ini dilaksanakan dengan pendekatan *Systematic Literature Review* (SLR) untuk mengidentifikasi, menganalisis, dan mensintesis tren serta inovasi dalam model dan algoritma yang digunakan pada analisis video kerumunan-padat.

Penyusunan SLR ini mengikuti proses penyusunan SLR yang diungkapkan oleh Xiao dan Watson [1]. Proses Xiao dan Watson dalam pembuatan SLR, terdiri dari tiga tahap utama (yaitu perencanaan peninjauan, pelaksanaan peninjauan, dan pelaporan peninjauan) dan delapan langkah proses umum, yang diilustrasikan pada Gambar 1.



Gambar 1 Proses Penyusunan SLR (Sumber Xiao dan Watson [1])

Langkah-langkah dalam pembuatan *Systematic Literature Review* (SLR) dimulai dengan merumuskan pertanyaan penelitian yang jelas dan terfokus (langkah *formulate the problem*). Selanjutnya, disusun protokol review yang menjelaskan tujuan, kriteria inklusi/eksklusi, strategi pencarian, serta prosedur analisis (langkah *develope and validate the review protocol*). Setelah itu dilakukan penelusuran literatur secara sistematis menggunakan kata kunci yang relevan di berbagai basis data (langkah *search the*

literature). Artikel yang ditemukan kemudian diseleksi berdasarkan judul, abstrak, dan teks penuh sesuai kriteria yang telah ditetapkan (langkah *screen for the inclusion*). Tahap berikutnya adalah menilai kualitas studi yang lolos seleksi untuk memastikan relevansi dan validitasnya (langkah *assess quality*). Data penting dari setiap artikel lalu diekstraksi menggunakan format standar (langkah *extracting data*), kemudian dianalisis dan disintesis untuk menemukan pola, kesenjangan, dan kontribusi ilmiah (langkah *analyzing and synthesizing data*). Terakhir, seluruh proses dan hasil review dilaporkan secara sistematis dalam bentuk narasi, tabel, dan visualisasi seperti diagram PRISMA (langkah *reporting the findings*).

2.1. Tujuan SLR

Naskah Penelitian ini dilaksanakan dengan pendekatan *Systematic Literature Review* (SLR) untuk mengidentifikasi, menganalisis, dan mensintesis tren serta inovasi dalam model dan algoritma yang digunakan pada analisis video kerumunan-padat. SLR ini menggunakan metode *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) guna menjamin transparansi dan keterulangan dalam proses peninjauan pustaka. Secara spesifik, SLR ini memiliki beberapa tujuan yang dirumuskan sebagai berikut:

1. Identifikasi metode dan algoritma yang digunakan dalam literatur.
2. Mengorganisasikan dan mengelompokkan model-model berdasarkan tujuan dan fungsi yang terdapat dalam literatur.
3. Analisis perkembangan tren dan ide yang ditemukan dalam literatur dari tahun 2010 hingga 2025.
4. Kaji penerapan teknologi, yang terdapat pada literatur, pada dunia nyata.
5. Identifikasi masalah/tantangan teknis dan kesenjangan penelitian yang belum terselesaikan.
6. Menyediakan landasan bagi peneliti selanjutnya dalam merancang pendekatan atau sistem baru yang lebih adaptif dan efisien.

2.2. Pertanyaan Penelitian (*Research Question*)

Format Dalam SLR ini, perumusan pertanyaan penelitian (*Research Questions / RQ*) dilakukan untuk membimbing proses pencarian, seleksi, dan analisis literatur secara sistematis. Pertanyaan-pertanyaan ini dirancang untuk mengarahkan fokus kajian pada

tren, teknik, serta gap riset dalam analisis video kerumunan padat. Adapun pertanyaan-pertanyaan penelitian yang diajukan dalam Tabel 1 berikut:

Tabel 1. Pertanyaan Penelitian yang diajukan dalam SLR ini.

Kode RQ	Pertanyaan Penelitian / <i>Research Question</i> (RQ)
RQ1	Apa saja model dan algoritma yang telah digunakan dalam analisis video kerumunan padat dalam literatur ilmiah?
RQ2	Bagaimana tren dan evolusi pendekatan tersebut berkembang dari tahun 2010 hingga 2025?
RQ3	Apa saja fokus aplikasi dan domain penerapan dari model dan algoritma yang diusulkan?
RQ4	Apa tantangan teknis, keterbatasan, dan celah penelitian (research gap) yang diidentifikasi dalam literatur?

2.3. Sumber dan Strategi Pencarian Literatur

SLR ini mengandalkan satu basis data utama, yaitu Scopus, karena merupakan salah satu indeks literatur ilmiah terkemuka yang mencakup jurnal-jurnal internasional bereputasi tinggi dan *peer-reviewed* dari berbagai disiplin ilmu yang relevan, termasuk ilmu komputer, teknik, sistem cerdas, dan pengolahan citra.

Strategi pencarian dikembangkan untuk menjangkau publikasi yang relevan dengan topik analisis video kerumunan-padat, dengan memadukan kata kunci utama dalam bahasa inggris dan diformulasikan dalam bentuk *Search Query* yang disajikan dalam Tabel 2 di bawah ini :

Tabel 2. Search Query untuk Scopus dan Jumlah Literatur yang diperoleh

Basis Data	Search Query	Jumlah Literatur Ilmiah
Scopus	video AND dense AND crowd	263

2.4. Kriteria Inklusi dan Eksklusi

SLR ini diharapkan hanya menganalisis publikasi yang relevan, terkini, dan berkualitas tinggi saja. Untuk itu kriteria inklusi dan eksklusi ditetapkan secara objektif, transparan, dan dapat direplikasi. Literatur ilmiah akan disertakan dalam analisis apabila memenuhi seluruh kriteria inklusi yang terdapat pada Tabel 3 di bawah ini:

Tabel 3. Kriteria Inklusi

No	Kriteria Inklusi	Penjelasan
I1	Publikasi tersedia di basis data Scopus	Menggaransi reputasi dan kualitas literatur
I2	Literatur berbahasa Inggris	Menjamin pemahaman dan konsistensi penilaian
I3	Diterbitkan dalam rentang tahun 2010–2025	Fokus pada tren dan inovasi dalam dekade terakhir
I4	Berupa literatur jurnal	Fokus pada literatur peer-reviewed dan kontribusi akademik
I5	Mempelajari analisis video kerumunan padat	Harus mencakup konteks crowd video secara eksplisit
I6	Mengusulkan atau mengevaluasi model/algoritma (misalnya: deteksi, pelacakan, prediksi, counting, anomaly)	Relevan terhadap fokus utama penelitian
I7	Tersedia teks lengkap (<i>full-text</i>) untuk dianalisis	Diperlukan untuk penilaian mendalam

Literatur ilmiah akan dikecualikan dari analisis apabila memenuhi salah satu dari kriteria eksklusi yang terdapat pada Tabel 4 di bawah ini:

Tabel 4. Kriteria Eksklusi

No	Kriteria Eksklusi	Penjelasan
E1	Bukan literatur ilmiah (misal: editorial, abstrak saja, book chapter, poster, whitepaper) atau prosiding konferensi	Tidak memiliki struktur ilmiah yang lengkap dan terverifikasi
E2	Fokus utamanya bukan pada crowd video analysis	Misalnya hanya membahas pedestrian tracking secara umum, bukan spesifik dalam kerumunan padat
E3	Literatur tidak menyajikan metode teknis (model/algoritma) secara eksplisit	Literatur bersifat konseptual atau ulasan umum tanpa kontribusi teknis konkrit
E4	Duplikasi (literatur yang sama dalam versi konferensi atau jurnal (pilih salah satu))	Hanya satu versi dengan konten paling lengkap yang dipilih
E5	Literatur dengan akses terbatas dan tidak dapat diakses secara penuh	Diperlukan untuk penilaian mendalam

Kriteria inklusi dan eksklusi dirancang untuk menjaga kesesuaian antara tujuan SLR dan ruang lingkup penelitian. Khususnya, ditekankan bahwa fokus harus pada analisis video kerumunan-padat, sehingga literatur yang membahas kerumunan dari perspektif sosial, statistik makro, atau non-visual tidak dimasukkan.

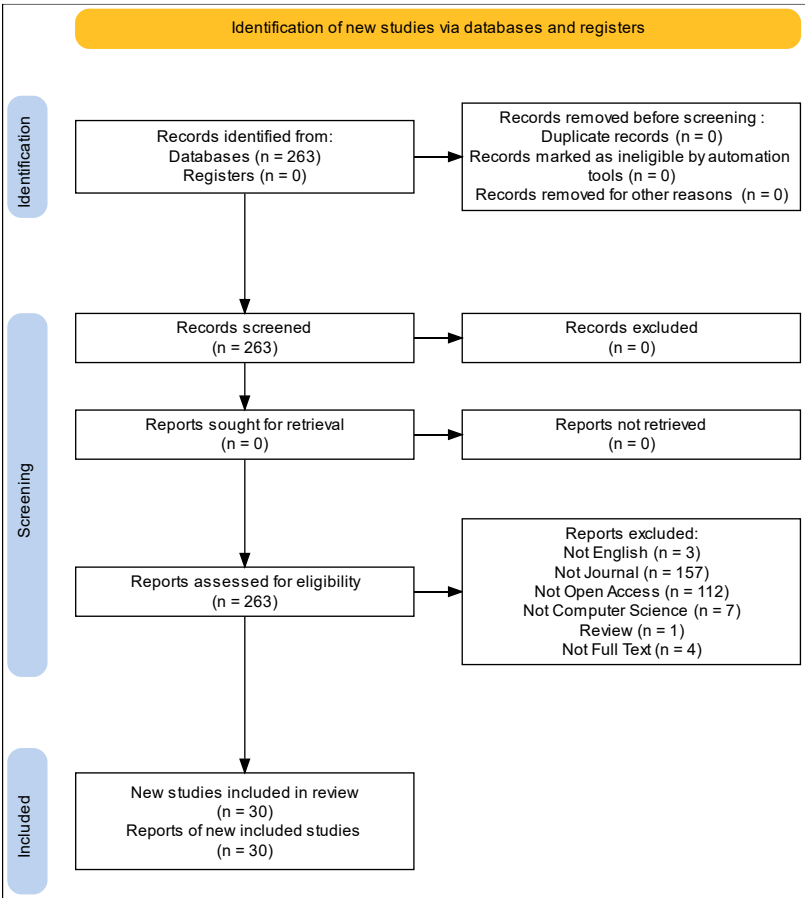
2.5. Proses Seleksi Literatur Ilmiah

Proses seleksi dilakukan secara sistematis dan berjenjang, mengikuti kerangka kerja *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA 2020). Proses seleksi terdiri dari empat tahap utama yang aktivitas dideskripsikan dalam Tabel 5

Tabel 5. Tahapan Seleksi Literatur Ilmiah

Tahap	Aktivitas	Deskripsi
1	Identifikasi	Mengumpulkan semua publikasi yang relevan dari basis data Scopus menggunakan strategi pencarian terstruktur
2	Penyaringan awal (<i>screening</i>)	Menghapus duplikasi dan memeriksa judul serta abstrak untuk mengevaluasi kesesuaian awal terhadap kriteria inklusi
3	Pemeriksaan kelayakan (<i>eligibility</i>)	Meninjau teks lengkap literatur untuk memastikan bahwa isi sesuai dengan kriteria inklusi dan eksklusi
4	Inklusi akhir	Literatur yang lolos tahap kelayakan dianalisis secara mendalam dan disertakan dalam sintesis literatur

Proses tahapan seleksi literatur ilmiah pada SLR ini disajikan dalam diagram PRISMA 2020 pada Gambar 2 menunjukkan alur metodologi untuk mencari literatur ilmiah terkait, memfilternya, dan menulis laporan berdasarkan fokus utama SLR ini. Tahapan seleksi dimulai dengan memasukan *search query* (yang tertulis pada Tabel 2) ke basis data Scopus, yang selanjutnya dilakukan proses *screening* sehingga tersisa 30 jurnal yang diteliti dalam SLR ini.



Gambar 2 Diagram PRISMA 2020 SLR

3. HASIL DAN PEMBAHASAN

SLR ini menggambarkan lanskap yang luas dan terus berkembang dalam analisis video kerumunan padat. Penelitian ini ditinjau dari segi model dan algoritma yang digunakan, perkembangan teknis dari tahun ke tahun, fokus aplikasi, serta tantangan dan celah penelitian yang masih ada. Bagian ini disusun dengan mengintegrasikan jawaban dari keempat pertanyaan penelitian (RQ1–RQ4).

3.1. Ragam Model dan Algoritma yang Digunakan (RQ1)

Berbagai pendekatan telah dikembangkan dalam analisis video kerumunan yang padat. Penelitian menunjukkan bahwa metode berbasis *deep learning* telah mendominasi dalam dekade terakhir. Arsitektur *Convolutional Neural Networks* (CNN) dan turunannya, seperti 3D ConvNet dan Transformer, menjadi pilihan utama untuk tugas-tugas seperti penghitungan kerumunan, deteksi anomali, dan pemodelan perilaku kerumunan. Sebagai contoh, Elmezain et al [2] menerapkan 3D ConvNet dan Multi-SVM untuk mendeteksi perilaku dalam video yang padat, sementara Cob-Parro et al [3] mengintegrasikan dense optical flow dengan model *deep learning* untuk mendeteksi kejadian *stampede* (bergerombol).

Pendekatan machine learning klasik, seperti *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN), masih digunakan, terutama dalam penelitian yang mengandalkan fitur visual sederhana atau kombinasi fitur yang dibuat secara manual.

Di sisi lain, model fisik seperti *Social Force Model* (SFM), *Hydrodynamic Model*, dan pendekatan simulatif (misalnya HyPedSim oleh Dang et al. [4]) tetap relevan dalam memodelkan interaksi antar individu dalam kerumunan yang padat.

Selain itu, terdapat juga pendekatan hibrida dan multimodal yang menggabungkan data visual dengan informasi spasial atau metadata lainnya. Beberapa penelitian mengombinasikan hasil dari drone, CCTV statis, kamera bergerak, dan data simulasi untuk menghasilkan hasil yang lebih *robust* dan adaptif.

Tiga puluh literatur ilmiah yang terpilih disajikan dalam bentuk tabel klasifikasi (Tabel 6) berdasarkan jenis pendekatan teknis, algoritma yang digunakan, serta sumber data video yang dianalisis.

Tabel 6. Klasifikasi 30 literatur ilmiah terpilih

No	Referensi (penulis & Tahun)	Kategori Pendekatan	Algoritma	Sumber Data Video
1	Aljuaid et al. (2023)[5]	<i>Deep Learning</i>	<i>Posture-based CNN, LSTM</i>	Simulasi & CCTV
2	Castellano et al. (2023)[6]	<i>Deep Learning</i>	<i>FCN + Clustering</i>	Drone
3	Cob-Parro et al. (2025)[3]	<i>Deep Learning</i>	<i>Dense Optical Flow + DL</i>	CCTV
4	Dang et al. (2024)[4]	Fisik/Simulatif	<i>HyPedSim Framework</i>	Simulasi
5	Dufour et al. (2025)[7]	Fisik/Simulatif	<i>Data Tracking (Ground Truth)</i>	CCTV
6	Elmezain et al. (2024)[2]	<i>Deep Learning + ML</i>	<i>3D ConvNet + Multi-SVM</i>	CCT
7	Fagette et al. (2014)[8]	<i>Machine Learning</i> klasik	<i>Texture + Clustering</i>	CCTV
8	Fan et al. (2018)[9]	Simulatif + Deep	<i>Crowd Navigation, LSTM</i>	Simulasi
9	Felemban et al. (2021)[10]	Deep Learning	<i>Trajectory-based CNN</i>	CCTV
10	Flagg & Rehg (2013)[11]	Simulatif	<i>Crowd Synthesis</i>	Simulasi
11	Hu et al. (2016)[12]	<i>Deep Learning</i>	<i>CNN Crowd Counting</i>	Gambar statis
12	Joshi & Patel (2025)[13]	<i>Deep Learning</i>	<i>Autoencoder, CNN</i>	CCTV
13	Li et al. (2019)[14]	<i>Deep Learning + Simulasi</i>	<i>DNN + Social Force Model</i>	Simulasi
14	Magdy et al. (2023)[15]	<i>Deep Learning</i>	<i>4D ConvNet</i>	CCTV
15	Mei et al. (2024)[16]	<i>Deep Learning</i>	<i>Global Collectiveness Metric</i>	Drone
16	Modi & Parikh (2022)[17]	Review (kompilasi)	Klasifikasi metode	N/A
17	Nishimura et al. (2022)[18]	<i>Deep Learning</i>	Transformer (ViewBirdiformer)	Ego-centric
18	Qi et al. (2023)[19]	<i>Deep Learning</i>	<i>Head Tracking + DL</i>	CCTV
19	Ren et al. (2022)[20]	<i>Deep Learning</i>	Multimodal Fusion	CCTV
20	Savitha & Ramesh (2019)[21]	<i>Machine Learning</i> klasik	<i>moDTA Model</i>	CCTV
21	Shah (2024)[22]	<i>Machine Learning</i> klasik	ML Classifier (SVM, DT)	Video Haji
22	Tomar et al. (2024)[23]	<i>Deep Learning</i>	PeopleNet	Kamera bergerak
23	Ullah et al. (2017)[24]	Simulatif	<i>Hydrodynamics Model</i>	Simulasi
24	Wang et al. (2018)[25]	Simulatif	<i>Direction-Collectiveness</i>	Simulasi
25	Washington et al. (2022)[26]	<i>Deep Learning + Annotasi</i>	<i>Video + Annotator CNN</i>	CCTV
26	Wong & Law (2023)[27]	<i>Deep Learning</i>	CCTV + Spatial Fusion	CCTV
27	Xu et al. (2022)[28]	<i>Deep Learning</i>	<i>Multi-object Tracking (MOT)</i>	CCTV
28	Xue et al. (2017)[29]	Simulatif	Lattice Boltzmann Model	Simulasi padat
29	Zhang et al. (2023)[30]	<i>Deep Learning</i>	<i>Multi-Scale Network + Depth</i>	CCTV
30	Zhu et al. (2021)[31]	<i>Federated Learning</i>	<i>FL + CNN Masked Face</i>	CCTV

3.2. Tren dan Evolusi Pendekatan (RQ2)

Bila melihat Tabel 7, maka terlihat pada tahun-tahun awal (2013–2016), pendekatan yang digunakan didominasi oleh metode tekstur, deteksi optik, dan model fisik. Namun, sejak tahun 2017, terjadi lonjakan publikasi yang didominasi oleh pendekatan *deep learning*.

Terlihat adanya pergeseran paradigma yang jelas dari metode konvensional menuju pendekatan berbasis kecerdasan buatan (AI). Sebagai contoh, penelitian yang dilakukan oleh Zhang et al.[30] mengusulkan adaptasi jaringan multi-skala berbasis *depth information* untuk penghitungan kerumunan, yang mencerminkan penerapan teknologi mutakhir dalam penghitungan kerumunan yang padat. Pada tahun 2025 mulai diadopsi teknik-teknik seperti *transfer learning*, *federated learning*, dan arsitektur berbasis transformer.

Tren ini juga ditandai dengan meningkatnya jumlah studi yang berfokus pada skenario ekstrem, seperti kerumunan dinamis, bencana, atau kondisi pencahayaan rendah, yang memerlukan pengembangan model yang lebih adaptif, efisien, dan mampu beroperasi secara real-time. Evolusi pendekatan dan model analisis video kerumunan padat dapat dilihat pada Tabel 7.

Tabel 7. Evolusi Pendekatan Model Analisis Kerumunan (2013-2025)

Tahun	Jumlah Jurnal	Pendekatan Dominan	Tren/Inovasi Terkait
2013	1	Simulatif	<i>Crowd synthesis</i> ; model statistik visual dasar
2014	1	<i>Machine Learning</i>	<i>Texture + clustering</i> ; <i>unsupervised learning</i> awal
2016	1	<i>Deep Learning</i>	<i>Crowd counting</i> berbasis CNN mulai dikenalkan
2017	2	Simulatif	<i>Hydrodynamics & Lattice Boltzmann model</i>
2018	2	Deep + Simulatif	LSTM untuk prediksi <i>crowd navigation</i>
2019	2	Deep + ML	Kombinasi DNN dengan simulasi klasik (SFM)
2021	1	<i>Deep Learning</i>	CNN untuk trajectory crowd movement
2021	3	<i>Deep Learning</i>	Transformer ViewBirdiformer, MOT (<i>multi-object tracking</i>), <i>spatial fusion</i>
2023	5	<i>Deep Learning</i>	4D ConvNet, multimodal, annotator CNN
2024	4	<i>Deep Learning</i>	PeopleNet, <i>Global collectiveness</i> , fused ConvNet
2025	6	Hybrid & FL	<i>Federated Learning</i> , <i>DeepPosture</i> , simulasi hybrid

Antara tahun 2013 dan 2025, telah terjadi pergeseran yang signifikan dari metode konvensional menuju pendekatan *deep learning* yang berbasis pada CNN, LSTM, dan Transformer. Penerapan *federated learning* serta teknik hybrid menandai suatu era baru dalam pemanfaatan analisis video kerumunan, khususnya dalam aplikasi nyata seperti kota pintar dan drone pengawas.

3.3. Fokus Aplikasi dan Domain Penerapan (RQ3)

Rincian mengenai fokus aplikasi dan domain penerapan dari literatur yang telah dipilih disajikan dalam Tabel 8 :

Tabel 8. Fokus Aplikasi dan Domain Penerapan

Product	Contoh Literatur	Tujuan Aplikasi	Teknologi Yang Digunakan
Keamanan Publik & Pemantauan CCTV	Wong & Law (2023)[27], Ren et al. (2022)[20], Shah (2024)[22]	Deteksi kemacetan, kepadatan tinggi, dan anomali	CNN, multimodal fusion, ensemble SVM
Manajemen Kerumunan Event Massal	Joshi & Patel (2025)[13], Qi et al. (2023)[19], Shah (2024)[22]	Prediksi lonjakan massa dan anomali saat haji, konser, demonstrasi	3D ConvNet, Head Tracking, Deep Classifier
Skenario Evakuasi dan Darurat	Li et al. (2019)[14], Fan et al. (2018)[9], Wang et al. (2018)[25]	Simulasi evakuasi, deteksi pelarian, navigasi robot dalam kerumunan	Deep+Simulatif (SFM, LSTM), Motion Prediction
Analisis Kepadatan & Estimasi Jumlah	Zhang et al. (2023)[30], Mei et al. (2024)[16], Hu et al. (2016)[12]	Menghitung jumlah orang, memetakan kepadatan spasial	CNN, Global Collectiveness Metric, Depth Fusion
Deteksi Kekerasan dan Risiko	Magdy et al. (2023)[15], Savitha & Ramesh (2019)[21]	Mendeteksi aksi kekerasan dan perilaku abnormal	4D CNN, MoDTA Approach
Drone & Kamera Bergerak	Castellano et al. (2023)[6], Tomar et al. (2024)[23]	Crowd tracking dari udara dan kamera bergerak (head-mounted)	FCN, Optical Flow, Moving-camera CNN
Simulasi & Crowd Synthesis	Dufour et al. (2025)[7], Flagg & Rehg (2013)[11], Dang et al. (2024)[4]	Menciptakan data sintetik dan menguji performa algoritma	Agent-based model, simulasi sosial, pedestrian dataset
Sistem Smart City / Urban Planning	Wong & Law (2023)[27], Washington et al. (2022)[26]	Integrasi pemantauan kerumunan ke dalam manajemen kota pintar	Integrasi spatial-GIS + CCTV, privacy-preserving ML
Asistensi Robotika & Navigasi	Fan et al. (2018)[9], Nishimura et al. (2022)[18]	Membantu robot menavigasi kerumunan tanpa tersesat	LSTM, Transformer-based View Recovery

Model dan algoritma yang diusulkan oleh literatur yang terpilih dapat diaplikasikan secara luas dalam berbagai domain operasional, misalnya :

- Pengawasan publik di ruang terbuka seperti stasiun, bandara, dan taman kota [27],
- Manajemen kerumunan dalam acara besar seperti haji [22], festival [7], dan konser,
- Perencanaan evakuasi dan deteksi kejadian darurat seperti *stampede* dan *panic escape* [25],
- *Sistem smart city* dan *surveillance* berbasis drone, yang menuntut model adaptif terhadap perspektif kamera yang berubah-ubah [6].

Hal ini mencerminkan kematangan teknis sekaligus tantangan praktis, di mana model tidak hanya harus akurat, tetapi juga efisien dan aman untuk digunakan dalam lingkungan yang padat manusia. Selain itu, tujuan dari RQ3 adalah untuk mengeksplorasi penerapan

praktis dari berbagai pendekatan analisis video yang telah diidentifikasi dalam literatur terkait kerumunan padat. Ini mencakup klasifikasi domain penggunaan, lingkungan penerapan, serta tingkat kesiapan dalam implementasi teknologi.

Pendekatan analisis video pada kerumunan padat telah mengalami perkembangan dari sekadar estimasi kepadatan menjadi sebuah sistem yang komprehensif untuk keamanan, navigasi, manajemen massa, dan pengambilan keputusan di perkotaan. Tren yang akan datang menunjukkan adanya integrasi yang kuat dengan pengawasan menggunakan drone, simulasi sosial, serta aplikasi kota pintar.

3.4. Tantangan, Keterbatasan, dan Celah Penelitian (RQ4)

Detail tantangan, keterbatasan, dan research gap yang ditemukan pada literatur terpilih disajikan dalam Tabel 9.

Tabel 9. Tantangan, Keterbatasan, dan Research Gap

Kategori Tantangan	Deskripsi	Contoh Literatur Terkait
1. Keterbatasan Dataset	- Minimnya dataset publik yang representatif untuk skenario crowd padat. - Keterbatasan anotasi manual dan data dari berbagai sudut pandang.	Dufour et al. (2025)[7], Zhang et al. (2023)[30], Washington et al. (2022)[26]
2. Kualitas Video dalam Kondisi Nyata	Penurunan performa model dalam kondisi gelap, berkabut, gerakan kamera, atau resolusi rendah.	Zhu et al. (2021)[31], Nishimura et al. (2022)[18]
3. Skalabilitas dan Efisiensi Komputasi	Model deep learning yang kompleks menuntut perangkat keras tinggi dan kurang cocok untuk inferensi real-time.	Magdy et al. (2023)[15], Qi et al. (2023)[19], Fan et al. (2018)[9]
4. Interpretabilitas dan Kepercayaan Model	Kinerja tinggi namun kurang dapat dijelaskan secara intuitif oleh pengguna awam atau pihak keamanan.	Joshi & Patel (2025)[13], Shah (2024)[22]
5. Adaptabilitas Model Lintas Skenario	Model yang dilatih pada satu skenario belum tentu efektif untuk lokasi berbeda karena perbedaan pola kerumunan.	Hu et al. (2016)[12], Savitha & Ramesh (2019)[21]
6. Tantangan Privasi dan Etika	Rekaman video kerumunan berisiko mengandung data sensitif (wajah, perilaku), menimbulkan kekhawatiran etik.	Washington et al. (2022)[26], Zhu et al. (2021)[31]
7. Kurangnya Validasi Lapangan	Banyak studi berbasis eksperimen laboratorium atau data simulasi, minim evaluasi dalam kondisi real-world.	Cob-Parro et al. (2025)[3], Dang et al. (2024)[4]

Selain itu, penelitian terkini masih jarang yang melakukan benchmarking lintas domain atau integrasi data multimodal secara sistematis, yang membuka peluang untuk pengembangan sistem yang lebih *robust* dan kontekstual.

4. KESIMPULAN

SLR ini menganalisis perkembangan model dan algoritma untuk analisis video kerumunan yang padat, yang dimulai dari tahun 2013 hingga 2025, dengan cara pendekatan *Systematic Literature Review* (SLR) terhadap 30 publikasi ilmiah. Tujuan dari SLR ini adalah memberikan pemahaman yang mendalam tentang evolusi teknis, tren penerapan, serta tantangan yang masih dihadapi di bidang analisis video kerumunan yang padat.

Hasil analisis menunjukkan bahwa pendekatan berbasis *deep learning* telah menjadi dominan dalam penelitian terkini, menggantikan metode konvensional seperti ekstraksi fitur visual manual dan model simulatif klasik. Arsitektur CNN, LSTM, dan Transformer berfungsi sebagai pilar utama dalam pengembangan sistem analisis kerumunan yang mampu melakukan berbagai tugas, termasuk pelacakan objek, deteksi anomali, estimasi kepadatan, dan prediksi perilaku. Selain itu, munculnya pendekatan model hibrida, multimodal, dan *federated learning* mencerminkan kecenderungan integrasi lintas teknologi untuk meningkatkan adaptabilitas dan privasi.

SLR ini juga menemukan bahwa aplikasi model-model tersebut semakin meluas ke berbagai domain, seperti keamanan publik, manajemen kerumunan massal, sistem *smart city*, pemantauan berbasis drone, dan skenario evakuasi darurat. Namun, terdapat tantangan nyata yang menghambat adopsi sistem secara luas di dunia nyata, termasuk keterbatasan dataset yang representatif, penurunan performa dalam kondisi ekstrem, kebutuhan komputasi yang tinggi, serta isu terkait interpretabilitas dan etika privasi.

SLR ini juga mengidentifikasi beberapa celah penelitian yang belum banyak dieksplorasi, seperti pengembangan model yang lebih ringan dan adaptif, validasi lintas skenario, serta eksplorasi penggunaan data non-visual atau sensor tambahan dalam konteks kerumunan.

Hasil penelitian ini diharapkan dapat menjadi referensi strategis bagi peneliti dan praktisi dalam merancang sistem analisis video kerumunan yang lebih akurat, *real-time*, dan kontekstual, serta mendorong inovasi yang dapat memperkuat keamanan dan efisiensi pengelolaan kerumunan di masa mendatang.

5. DAFTAR PUSTAKA

- [1] Y. Xiao and M. Watson, "Guidance on Conducting a Systematic Literature Review," Mar. 01, 2017, *SAGE Publications Inc.* doi: 10.1177/0739456X17723971.
- [2] M. Elmezain, A. S. Maklad, M. Alwateer, M. Farsi, and H. M. Ibrahim, "Analyzing Crowd Behavior in Highly Dense Crowd Videos Using 3D ConvNet and Multi-SVM," *Electronics (Switzerland)*, vol. 13, no. 24, Dec. 2024, doi: 10.3390/electronics13244925.

- [3] A. C. Cob-Parro, C. Losada-Gutiérrez, and M. Marrón-Romera, “Stampede detector based on deep learning models using dense optical flow,” *Eng Appl Artif Intell*, vol. 142, Feb. 2025, doi: 10.1016/j.engappai.2024.109940.
- [4] H. T. Dang, B. Gaudou, and N. Verstaevael, “HyPedSim: A Multi-Level Crowd-Simulation Framework—Methodology, Calibration, and Validation †,” *Sensors*, vol. 24, no. 5, Mar. 2024, doi: 10.3390/s24051639.
- [5] H. Aljuaid *et al.*, “Postures anomaly tracking and prediction learning model over crowd data analytics,” *PeerJ Comput Sci*, vol. 9, 2023, doi: 10.7717/peerj-cs.1355.
- [6] G. Castellano, E. Cotardo, C. Mencar, and G. Vessio, “Density-based clustering with fully-convolutional networks for crowd flow detection from drones,” *Neurocomputing*, vol. 526, pp. 169–179, Mar. 2023, doi: 10.1016/j.neucom.2023.01.059.
- [7] O. Dufour *et al.*, “Dense Crowd Dynamics and Pedestrian Trajectories: A Multiscale Field Dataset from the Festival of Lights in Lyon,” *Scientific Data*, vol. 12, no. 1, Dec. 2025, doi: 10.1038/s41597-025-04732-3.
- [8] A. Fagette, N. Courty, D. Racoceanu, and J. Y. Dufour, “Unsupervised dense crowd detection by multiscale texture analysis,” *Pattern Recognit Lett*, vol. 44, pp. 126–133, Jul. 2014, doi: 10.1016/j.patrec.2013.09.020.
- [9] T. Fan *et al.*, “Getting Robots Unfrozen and Unlost in Dense Pedestrian Crowds,” Sep. 2018, [Online]. Available: <http://arxiv.org/abs/1810.00352>
- [10] E. Felemban, S. D. Khan, A. Naseer, F. U. Rehman, and S. Basalamah, “Deep Trajectory Classification Model for Congestion Detection in Human Crowds,” *Computers, Materials and Continua*, vol. 68, no. 1, pp. 705–725, Mar. 2021, doi: 10.32604/cmc.2021.015085.
- [11] M. Flagg and J. M. Rehg, “Video-based crowd synthesis,” *IEEE Trans Vis Comput Graph*, vol. 19, no. 11, pp. 1935–1947, 2013, doi: 10.1109/TVCG.2012.317.
- [12] Y. Hu, H. Chang, F. Nian, Y. Wang, and T. Li, “Dense crowd counting from still images with convolutional neural networks,” *J Vis Commun Image Represent*, vol. 38, pp. 530–539, Jul. 2016, doi: 10.1016/j.jvcir.2016.03.021.

- [13] K. V. Joshi and N. M. Patel, "Supervised Deep Learning Approaches for Anomaly Detection and Recognition in Crowd Scenes," *Electronic Letters on Computer Vision and Image Analysis*, vol. 24, no. 1, pp. 31–50, 2025, doi: 10.5565/REV/ELCVIA.1631.
- [14] X. Li, Y. Liang, M. Zhao, C. Wang, H. Bai, and Y. Jiang, "Simulation of evacuating crowd based on deep learning and social force model," *IEEE Access*, vol. 7, pp. 155361–155371, 2019, doi: 10.1109/ACCESS.2019.2949106.
- [15] M. Magdy, M. W. Fakhr, and F. A. Maghraby, "Violence 4D: Violence detection in surveillance using 4D convolutional neural networks," *IET Computer Vision*, vol. 17, no. 3, pp. 282–294, Apr. 2023, doi: 10.1049/cvi2.12162.
- [16] L. Mei, M. Yu, L. Jia, and M. Fu, "Crowd Density Estimation via Global Crowd Collectiveness Metric," *Drones*, vol. 8, no. 11, Nov. 2024, doi: 10.3390/drones8110616.
- [17] H. S. Modi and D. A. Parikh, "A Survey on Crowd Anomaly Detection," *International Journal of Computing and Digital Systems*, vol. 12, no. 4, pp. 1081–1096, Oct. 2022, doi: 10.12785/ijcds/120187.
- [18] M. Nishimura, S. Nobuhara, and K. Nishino, "ViewBirdiformer: Learning to recover ground-plane crowd trajectories and ego-motion from a single ego-centric view," Oct. 2022, [Online]. Available: <http://arxiv.org/abs/2210.06332>
- [19] Z. Qi, M. Zhou, G. Zhu, and Y. Xue, "Multiple Pedestrian Tracking in Dense Crowds Combined with Head Tracking," *Applied Sciences (Switzerland)*, vol. 13, no. 1, Jan. 2023, doi: 10.3390/app13010440.
- [20] G. Ren, X. Lu, and Y. Li, "Research on 24-Hour Dense Crowd Counting and Object Detection System Based on Multimodal Image Optimization Feature Fusion," *Sci Program*, vol. 2022, 2022, doi: 10.1155/2022/9863066.
- [21] C. Savitha and D. Ramesh, "Crowd behavior analysis using moDTA approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 1, pp. 484–494, Jul. 2019, doi: 10.11591/ijeecs.v15.i1.pp484-494.
- [22] A. A. Shah, "A Machine Learning Model for Crowd Density Classification in Hajj Video Frames," 2024. [Online]. Available: www.ijacsa.thesai.org
- [23] A. Tomar, S. Kumar, and B. Pant, "PeopleNet: A Novel People Counting Framework for Head-Mounted Moving Camera Videos," *International Journal of*

- Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 6, pp. 61–73, 2024, doi: 10.9781/ijimai.2023.04.002.
- [24] H. Ullah, M. Uzair, M. Ullah, A. Khan, A. Ahmad, and W. Khan, “Density independent hydrodynamics model for crowd coherency detection,” *Neurocomputing*, vol. 242, pp. 28–39, Jun. 2017, doi: 10.1016/j.neucom.2017.02.023.
- [25] M. Wang, F. Chang, and Y. Zhang, “Crowd escape event detection based on direction-collectiveness model,” *KSII Transactions on Internet and Information Systems*, vol. 12, no. 9, pp. 4355–4374, Sep. 2018, doi: 10.3837/tiis.2018.09.013.
- [26] P. Washington *et al.*, “Crowd annotations can approximate clinical autism impressions from short home videos with privacy protections,” *Intell Based Med*, vol. 6, Jan. 2022, doi: 10.1016/j.ibmed.2022.100056.
- [27] V. W. H. Wong and K. H. Law, “Fusion of CCTV Video and Spatial Information for Automated Crowd Congestion Monitoring in Public Urban Spaces,” *Algorithms*, vol. 16, no. 3, Mar. 2023, doi: 10.3390/a16030154.
- [28] B. Xu, D. Liang, L. Li, R. Quan, and M. Zhang, “An Effectively Finite-Tailed Updating for Multiple Object Tracking in Crowd Scenes,” *Applied Sciences (Switzerland)*, vol. 12, no. 3, Feb. 2022, doi: 10.3390/app12031061.
- [29] Y. Xue, P. Liu, Y. Tao, and X. Tang, “Abnormal prediction of dense crowd videos by a purpose-driven lattice Boltzmann model,” *International Journal of Applied Mathematics and Computer Science*, vol. 27, no. 1, pp. 181–194, Mar. 2017, doi: 10.1515/amcs-2017-0013.
- [30] P. Zhang, W. Lei, X. Zhao, L. Dong, and Z. Lin, “An Adaptive Multi-Scale Network Based on Depth Information for Crowd Counting †,” *Sensors*, vol. 23, no. 18, Sep. 2023, doi: 10.3390/s23187805.
- [31] R. Zhu, K. Yin, H. Xiong, H. Tang, and G. Yin, “Masked Face Detection Algorithm in the Dense Crowd Based on Federated Learning,” *Wirel Commun Mob Comput*, vol. 2021, 2021, doi: 10.1155/2021/8586016.