PENGIRAAN HIMPUNAN ORANG RAMAI BERDASARKAN PEMBELAJARAAN MENDALAM

Leong Zi Ying

Kok Ven Jyn

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Analisis himpunan orang ramai adalah topik penyelidikan yang penting untuk memahami sifat dan dinamik individu dalam tempat yang sesak. Analisis himpunan orang ramai sangat penting untuk pengurusan himpunan orang ramai terutamanya di kawasan awam, contohnya reka bentuk ruang awam atau persekitaran maya. Terdapat batasan ruang berlaku apabila perhimpunan orang ramai dengan jumlah populasi manusia yang semakin meningkat. Hal ini seterusnya boleh menyebabkan kesesakan berlaku jika tidak diuruskan dengan baik. Oleh itu, faktor kepadatan orang ramai dalam sistem pengawasan perlu diurus dengan baik. Namun begitu, penyelidikan sebelumnya menunjukkan bahawa sukar untuk menganggarkan jumlah orang ramai kerana variasi skala individu dalam himpunan orang ramai disebabkan oleh gangguan perspektif. Hal ini demikian, projek ini bertujuan untuk membangunkan algoritma yang dapat mengekstrak ciri-ciri yang berkaitan dengan skala supaya mengatasi variasi skala. Model yang dicadangkan dapat mengira jumlah individu yang terdapat dalam himpunan padat. Untuk mengatasi masalah variasi skala, projek ini mengusulkan kerangka berdasarkan Rangkaian Neural Konvolusional untuk mengira orang ramai. Model yang dicadangkan ini menerapkan VGG-16 sebagai tulang belakang dengan menggabungkan rangkaian tumpuan dan konvolusi pendilatan piramid yang menggunakan kernel pendilatan untuk memperbesar medan penerimaan dan menggantikan operasi “pooling”. Akhirnya, peta ciri akan dihasilkan untuk menganggarkan bilangan orang dalam gambar. Peta kepadatan dengan jumlah himpunan orang ramai akan digunakan sebagai penilaian bagi model ini. Eksperimen yang lanjut akan dilakukan dengan set data penanda aras awam seperti set data ShanghaiTech Bahagian A and Bahagian B. https://github.com/ziying23/Dense-Crowd-Counting-via-Deep-Learning.git

1 PENGENALAN

Himpunan orang ramai merujuk kepada individu dari set yang sama atau berbeza yang berada dalam satu kumpulan. Analisis orang ramai merupakan topik kajian yang penting untuk memahami tingkah laku dan pergerakan individu dalam himpunan orang ramai. Analisis himpunan orang ramai sangat penting untuk pengurusan himpunan

orang ramai terutamanya di kawasan awam, contohnya reka bentuk ruang awam atau persekitaran maya. Dengan populasi yang semakin meningkat, ruang menjadi terhad dan boleh menyebabkan kesesakan orang ramai jika tidak diuruskan dengan baik. Sebagai contohnya, tragedi 'Hajj Stampede' yang berlaku pada tahun 2015, lebih daripada 2000 orang warga Jemaah telah mengalami kematian disebabkan kesesakan (R. Kasolowsky 2015). Hal ini telah menyedari masyarakat tentang kepentingan pengurusan himpunan orang ramai terutamanya di kawasan tumpuan seperti tumpuan pelancongan dan bandaraya.

Maka, faktor kepadatan orang ramai dalam sistem pengawasan perlu diurus dengan baik. Selain itu, mekanisma pengiraan orang ramai boleh diimplementasi untuk mengira peserta dalam aktiviti protes atau program tertentu. Penganggaran ini dapat digunakan untuk mengawal pengaliran orang ramai bagi mengelakkan kesesakan yang melampau, penyerbuan dan merancang laluan pemindahan mangsa bencana. Di samping itu, penjarakan sosial (TPB Thu et al. 2020) merupakan satu cara yang berkesan untuk mengawal penularan virus semasa pandemik Covid-19 yang berlaku pada masa ini. Kesimpulannya, pengiraan oang ramai perlu dilakukan untuk mengawasi pergerakan orang ramai pada waktu sebenar.

Walaupun banyak usaha dan kemajuan telah dilakukan dalam penyelidikan, tetapi analisis himpunan orang ramai tetap merupakan cabaran yang besar kerana permasalahan gangguan perspektif, kekacauan latar belakang dan pencahayaan kompleks yang boleh menyebabkan kesalahan anggaran kepadatan orang ramai. Kegagalan pendekatan tradisional seperti kaedah pengesanan dan regresi telah mewujudkan pendekatan moden iaitu algoritma berdasarkan Rangkaian Neural Konvolusional (CNN) kerana mencapai kejayaan dalam penghitungan himpunan orang ramai. Hasil kerja sebelumnya menunjukkan bahawa kejayaan pendekatan berasaskan CNN untuk meramal peta kepadatan dengan klasifikasi dan pengiktirafan. Suatu tinjauan (V. A. Sindagi & V. M. Patel 2018) telah dilakukan untuk menganalisis pelbagai kaedah penghitungan orang ramai berdasarkan CNN. Masalah kesamaan penampilan tinggi dan perubahan perspektif yang kompeks menyebabkan kaedah berasaskan CNN tidak dapat berfungsi dengan baik. Oleh itu, projek ini memberi tumpuan kepada masalah variasi skala. Masalah variasi skala berlaku disebabkan oleh kepelbagaian jarak antara kamera dan orang ramai tersebut. Ukuran objek berkadar

songsang dengan jarak dari kamera. Perspektif dan jarak kamera dengan masing-masing objek boleh menyebabkan ukuran berbeza dalam gambar yang diambil. Oleh itu, model yang dicadangkan ini harus mengatasi masalah ini untuk menjalankan pengiraan himpunan orang ramai dari gambar.

2 PENYATAAN MASALAH

Kebelakangan ini, Rangakaian Neural Konvolusional (CNN) berdasarkan penglihatan komputer disarankan kerana pencapaian yang lebih baik berbanding dengan pendekatan tradisional. Walaupun CNN menunjukkan prestasi yang baik dalam penggiraan himpunan orang ramai, namun pendekatan ini juga diganggu oleh pembentukkan variasi skala yang besar dalam gambar. H. Bai, S. Wen dan S. Chan (2019) mendakwa bahawa variasi terbentuk kerana perspektif dan jarak kamera terhadap objek, kedua-dua objek muncul dalam ukuran yang berbeza dalam gambar walaupun mereka memiliki ukuran fizikal yang sama secara nyata (Rajah 1). Dalam tempat yang padat dengan orang ramai, individu mungkin menghadapi variasi skala besar yang boleh mempengaruhi ketepatan pengiraan orang ramai.

Secara kesimpulannya:

- Kesukaran dan tidak tepat dalam menganggarkan jumlah orang dari gambar himpunan orang ramai disebabkan oleh variasi skala.
- Model yang sedia ada memberi ketepatan yang rendah dalam himpunan orang ramai yang sangat padat.

Rajah 1: Gambar dari ShanghaiTech Part B.

Sumber: Set data ShanghaiTech Part_B

Rajah 1 merujukkan orang serupa kelihatan berbeza secara visual apabila mereka tidak berada dari jarak atau sudut yang sama dari kamera dan orang yang sama dapat kelihatan kecil dalam gambar, tetapi jauh lebih besar pada yang lain.

3 OBJEKTIF KAJIAN

Objektif utama projek ini bertujuan untuk mengbangunkan algoritma yang dapat melakukan pengiraan himpunan orang ramai. Eksperimen ekstensif akan dilakukan dengan menggunakan set data penanda aras awam seperti set data. Objektif projek ini adalah seperti berikut:

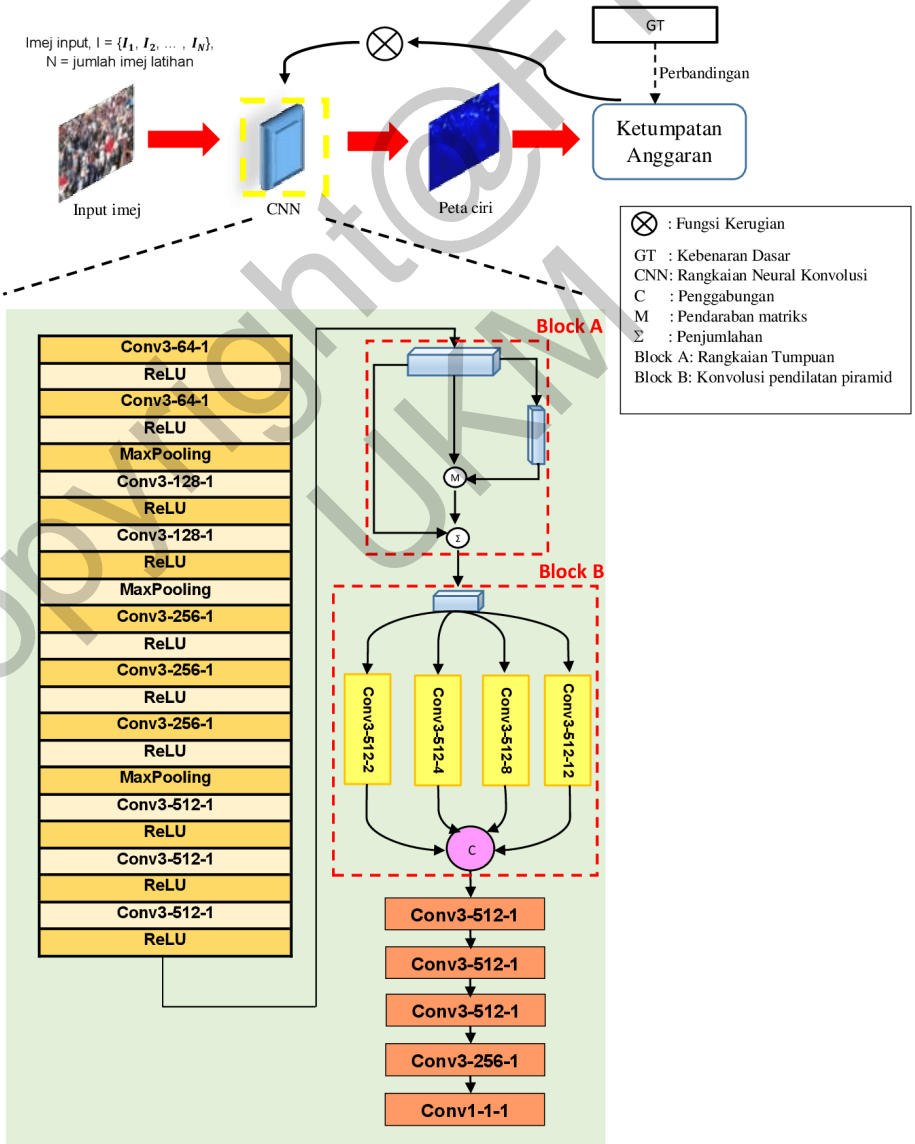
a) Mengkaji algorithma yang mampu mengekstrak ciri-ciri yang berbeza dari latar depan dan latar belakang.
b) Mengkaji model yang skala tetap (scale-invariant) dengan kernel dilebarkan untuk anggaran kepadatan himpunan orang ramai.
c) Membanding pendekatan yang sedia ada dengan set data ShanghaiTech Bahagian A dan Bahagian B.

4 METHOD KAJIAN

4.1 Seni Bina Model

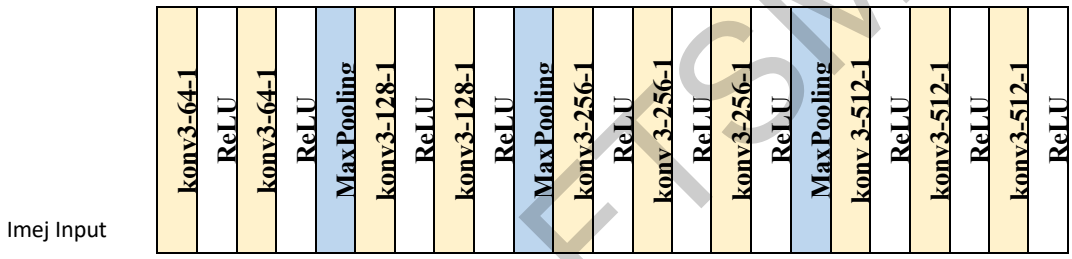
Kajian sebelumnya menunjukkan pencapaian yang berjaya dalam pengiraan himpunan orang ramai, maka model kami menerapkan jaringan rangka kerja rangkaian konvolusional untuk mempelajari ciri-ciri dari gambar input dan menghasilkan estimasi kepadatan pengiraan orang ramai. Merujuk kepada idea CSRNet (Y. Li, X. Zhang & D. Chan 2018), model yang dicadangkan mengekalkan sepuluh lapisan pertama VGG-16 dengan hanya tiga lapisan “pooling” sebagai tulang belakang model cadangan. VGG-16 menunjukkan pencapaian yang baik dalam pembelajaran pemindahan *(transfer learning)* (S. Liu & W. Deng. 2015) dan ia mampu mengekstrak ciri (features) sederhana hingga kompleks. Tujuan menggunakan sepuluh lapisan pertama dengan tiga lapisan “pooling” dan membuang lapisan “fully-connected” (FC) dalam VGG-16 adalah untuk mengurangkan kesan buruk terhadap ketepatan output yang disebabkan

oleh operasi “pooling”. Saiz output dari rangkaian VGG-16 adalah 1/8 daripada saiz input asal. Oleh kerana timbunan lapisan konvolusional dan “pooling” yang berterusan mengurangkan taburan ruang (spatial distribution) yang menjadikan peta ciri buruk (Brownlee & Jason. 2019), model yang dicadangkan menggunakan lapisan konvolusasi pendilatan untuk mengekstrak maklumat yang lebih mendalam tetapi mengekalkan resolusi output. Selain itu, mekanisme perhatian digabungkan untuk menggabungkan ciri-ciri pelbagai skala dan mengurangkan kekacauan latar belakang. Oleh kerana model cadangan (Rajah 2) menerapkan rangkaian tumpuan dan konvolusi dilatsi piramid, secara ringkas nama model boleh didefinisikan sebagai “Attention Pyramid Dilated Network” (APDNet).



Rajah 2: Reka bentuk keseluruhan model cadangan, APDNet. Imej input disalurkan ke dalam VGG-16 untuk pengekstrakan ciri kemudian disalurkan ke lapisan rangkaian tumpuan dan seterusnya kepada konvolusi pendilatan piramid. Parameter lapisan konvolusional dilambang sebagai "konv-(saiz kernel)-(jumlah penapis)-(kadar pendilatan)". Peta ketumpatan anggaran dihasilkan sebagai output.

4.2 Rangkaian pra-terlatih VGG-16 *(VGG-16 pre-trained network)*



Rajah 3: Empat set konvolusi pertama dalam rangkaian VGG-16. Parameter lapisan konvolusional dilambangkan sebagai "konv-(saiz kernel)-(jumlah penapis)-(kadar pendilatan)".

Pertama sekali, set data yang akan digunakan sebagai input dalam model ini adalah gambar himpunan orang ramai yang terdiri daripada gambar RGB yang mempunyai tiga saluran warna C = 3. Proses dalam APDNet bermula dengan pengambilan gambar dengan input $I \in \mathbb{R}^{H \times W \times C}$. Untuk memastikan prestasi model, gambar dipotong kepada 9 tampalan *(patches)* dari setiap gambar di lokasi yang berlainan, sedangkan setiap tampalan adalah ukuran 1/4 dari gambar asal. 4 tampalan pertama tidak bertindih dan merupakan empat perempat daripada gambar asal, sementara selebihnya lima tampalan dipotong secara rawak. Untuk meningkatkan saiz data latihan, model yang dicadangkan mencerminkan tampalannya. Seterusnya, setiap tampalan akan diubah saiz kepada 256 × 256. Y. Zhang et al. (2016) menyatakan bahawa terlalu banyak persampelan rendah *(downsampling)* akan menghasilkan peta ciri yang lebih kecil dan resolusi yang lebih rendah.

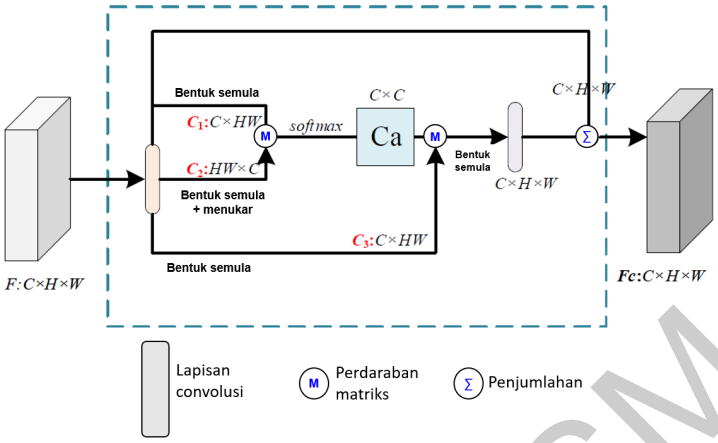
Oleh itu, untuk memperolehi peta kepadatan orang ramai yang berkualiti tinggi, model ini menggunakan 13 lapisan pertama dari VGG-16 untuk mengekstrak peta ciri pelbagai skala *(multi-scale)*. 13 lapisan terdiri daripada 10 lapisan konvolusi dan 3 lapisan “max-pooling”. Rangkaian pra-terlatih VGG-16 adalah telah diubahsuai dengan

membuang lapisan “fully-connected” (FC) kerana lapisan tersebut bertujuan untuk klasifikasi manakala projek ini hanya memerlukan regresi. “Rectified Linear Unit” (ReLU) digunakan sebagai fungsi pengaktifan. Rangkaian ini dibina dengan menyusun berulang kali lapisan konvolusional dengan kernel 3 x 3 dan lapisan “max-pooling” dengan penapis 2 x 2. Saiz output VGG-16 dalam model ini adalah 1/8 dari saiz input asal kemudian diteruskan ke rangkaian tumpuan.

4.3 Block A: Rangkaian tumpuan *(Attention Network)*

Saiz output VGG-16 adalah 1/8 dari saiz input asal dan dihantar ke lapisan rangkaian tumpuan ini. Rangkaian tumpuan dicadangkan adalah kerana kesusahan untuk membezakan latar depan dan belakang dalam himpunan orang ramai yang padat. Tetapi rangkaian perhatian dapat menyelesaikan masalah ini (Sanping Zhou et al. 2019). Rajah 4 menunjukkan seni bina rangkaian tumpuan. Secara konkrit, bagi peta ciri tertentu, F $\in \mathbb{R}^{C \times H \times W}$, mana C mewakili bilangan saluran, dan H mewakili ketinggian serta H mewakili kelebaran. Pertama sekali, satu lapisan 1 x 1 dilaksanakan dan kemudian membentuk semula *(reshaping)* atau menukar *(transpose)*, dua peta ciri akan dibentuk iaitu C_1 dan C_2. Untuk menghasilkan peta perhatian, matriks opera si pendaraban (matrix multiplication) dan “softmax” digunakan pada C_1 dan C_2. Saiz output daripada rangkaian tumpuan adalah 1/10 dari saiz input asal kepada menggunakan lapisan “average pooling”. Secara ringkas, prosesnya dapat dirumuskan seperti berikut, C_a^{ji} mewakil saluran yang dipengaruhi:

$$C_a^{ji} = \frac{\exp(C_1^i * C_2^j)}{\sum_{i=1}^{C}(C_1^i * C_2^j)} \quad \ldots(3)$$



Rajah 4: Perincian rangkaian tumpuan.

4.4 Block B: Konvolusi pendilatan piramid *(Pyramid Dilated Convolution)*

Lapisan "pooling" seperti "max-pooling" dan "average pooling" mengurangkan taburan ruang yang menyebabkan peta ciri menjadi kualiti rendah (Brownlee & Jason 2019). Oleh itu, model ini tidak terus menimbun lebih banyak lapisan konvolusi dan lapisan "pooling". APDNet melaksanakan konvolusi pendilatan untuk mengekalkan resolusi output. $x(l, w)$ adalah input konvolusi pendilatan manakala $y(l, w)$ adalah output konvolusi pendilatan dan $p(i, j)$ mewakili lapisan dengan panjang *L* and lebar *W*. Kadar pendilatan ditunjukkan sebagai parameter *r*. Konvolusi pendilatan 2 dimensi dapat didefinisikan sebagai:

$$y(l, w) = \sum_{i=0}^{L} \sum_{j=1}^{w} x\,(l + r * i, w + r * j)p(i, j) \quad \text{... (4)}$$

APDNet menyambungkan output dari rangkaian tumpuan iaitu peta ciri yang bersaiz 1/10 dari imej asal. Untuk mengatasi skala kepala yang bervariasi dalam imej, model yang diusulkan menerapkan konvolusi pendilatan untuk meningkatkan medan penerimaan *(receptive field)* dan saiz kernel yang berbeza untuk mempelajari saiz skala yang berbeza. Antara kadar pendilatan yang digunakan ialah 2, 4, 8 dan 12 yang mana dicadangkan (Xinya Chen et al. 2019). Rajah 5 menunjukkan konvolusi pendilatan piramid ini terdiri daripada empat lapisan konvolusi pendilatan yang selari. Konvolusi pendilatan adalah konvolusi dengan lubang seperti yang digambarkan dalam Rajah 5. Tujuan menggunakan pendilatan adalah untuk memperbesarkan medan penerimaan tanpa kehilangan resolusi peta ciri. Tambahan pula, kaedah ini tidak ada parameter tambahan, menjadikannya penyelesaian terbaik untuk pengiraan himpunan orang ramai. Hal ini demikian, APDNet dapat menangkap ciri pelbagai skala dan invarian dalam variasi skala imej. Dengan ini, APDNet dapat mudah menyesuaikan diri dengan pemandangan pelbagai skala dengan struktur yang lebih sederhana dan kurang kerumitan latihan. Untuk menjamin saiz output imej sama dengan saiz input iaitu 1/10 dari imej asal, model cadangan menggunakan "padding" untuk menambah piksel dalam gambar. Persamaan 5 menunjukkan cara untuk mendapat nilai "padding" bagi kadar pendilatan yang berbebeza. Contohnya, kadar pendilatan 2 menggunakan "padding" 2,

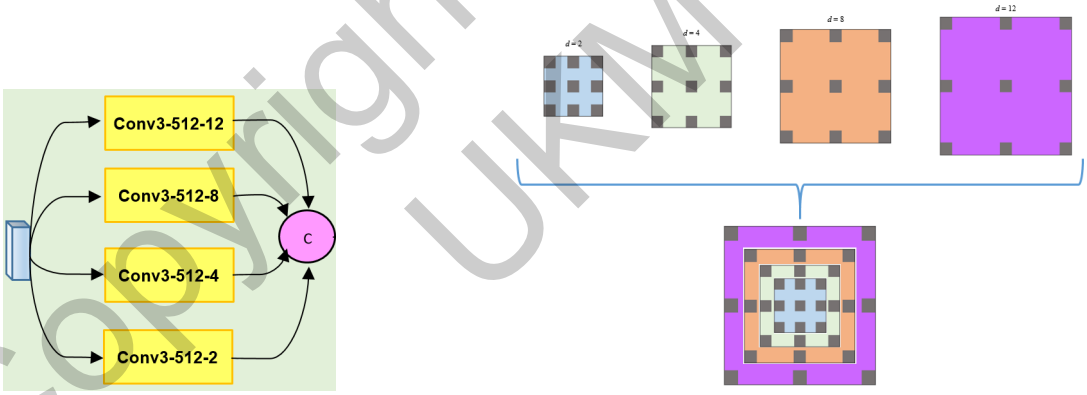
kadar pendilatan 4 menggunakan “padding” 4, kadar pendilatan 8 menggunakan “padding” 8 dan kadar pendilatan 12 menggunakan “padding” 12. Kaedah ini yang menyebabkan saiz imej output menjadi sama dengan saiz imej input ke konvolusi pendilatan piramid. Pada akhirnya, peta ciri tersebut diintegrasi bagi pengiraan himpunan oarang ramai.

$$p \approx \frac{k + (k-1)(d-1)}{2} \quad \ldots(5)$$

di mana $p = \ saiz\ padding,$

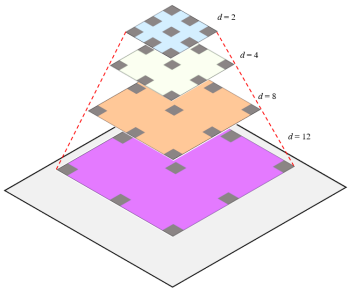
$$k = saiz\ kernel,$$

$$d = \ kadar\ pendilatan$$



Rajah 5: Pemandangan 2-dimensi konvousi pendilatan piramid (d = kadar pendilatan).

Rajah 5 menunjukkan empat lapisan konvolusi pendilatan yang berbeza dapat bergabung dengan saiz “padding” yang berbeza dengan merujuk kepada persamaan 5 supaya menghasilkan peta ciri yang bersaiz sama. Dengan “padding” yang sesuai, empat piksel numerik selari iaitu dari kadar pendilatan 2, 4, 8 dan 12 dapat digabungkan bersama dengan gabungan *(concatenation)*. Oleh itu, ia membentuk bentuk piramid untuk empat lapisan konvolusi pendilatan yang selari ini seperti Rajah 6.



Rajah 6: Pemandangan 3-dimensi konvousi pendilatan piramid (d = kadar pendilatan).

4.5 Pengoptimuman

Kami mengoptimumkan model tersebut dengan penurunan kecerunan stokastik yang diterapkan dengan kadar pembelajaran 1e-6 semasa latihan. Model ini dapat dilatih secara langsung sebagai struktur sempurna. Untuk memastikan prestasi model, kami juga menggunakan fungsi kerugian untuk mengukur perbezaan antara peta kepadatan yang dianggar dan kebenaran dasar.

4.6 Fungsi Kerugian

Fungsi kerugian adalah kaedah menilai sejauh mana algoritma memodelkan set data. Semakin rendah output fungsi kerugian, semakin baik prestasi model. Untuk mengira peta kepadatan dan kebenaran dasar, kebanyakan kerja yang ada (L. Zeng et al. 2017) menggunakan jarak Euclidean untuk melatih rangkaian tersebut. Oleh itu, model yang dicadangkan melaksanakan fungsi kerugian ini untuk mengurangkan kesalahan dalam ramalan. Dalam Persamaan 6, N mewakili jumlah gambar latihan, Θ adalah parameter model, sementara X_i adalah gambar input dan F_i adalah kebenaran dasar X_i. Akhir sekali, $F(X_i;\ \theta)$ adalah peta ketumpatan anggaran yang dihasilkan oleh model. Fungi kerugian ditakrifkan seperti berikut:

$$L(\theta) = \frac{1}{2N}\sum_{i=1}^{N} ||F(X_i;\ \theta) - F_i||_2^2 \qquad \ldots(6)$$

5 HASIL KAJIAN

5.1 Pengujian

Bagi tahap pengujian, model yang akan diuji dengan kajian sebelumnya adalah seperti berikut:

Model Ujian	Spesifikasi
Baseline	*Baseline* terdiri daripda rangkaian VGG-16 yang telah diubahsuai seperti di Bahagian 4.3.1.
***Baseline* + Konvolusi pendilatan piramid**	Konvolusi pendilatan piramid ditambah ke *Baseline* tersebut.
***Baseline* + Rangkaian tumpuan + Konvolusi pendilatan piramid**	Model candangan, iaitu "Attention Pyramid Dilated Network" (APDNet) dengan gabungan *Baseline*, rangkaian tumpuan dan konvolusi pendilatan piramid.

Jadual 1: Ringkasan model ujian.

5.2 Kaedah penilai

Merujuk kepada karya sebelumnya (L. Zeng et al. 2017) (Y. Li, X. Zhang & D. Chan 2018), metrik penilaian untuk pengiraan himpunan orang ramai diukur oleh dua metrik iaitu "Mean Absolute Error" (MAE) dan "Mean Square Error" (MSE). MAE mengukur magnitud rata-rata kesalahan dalam sekumpulan ramalan dan MSE adalah punca kuasa dua purata perbezaan kuasa dua antara ramalan dan pemerhatian sebenar, GT (C. J. Willmott & K Matsuura 2019). Persamaan (6) (7) ditaktifkan seperti di bawah, di mana *NT* ialah bilangan gambar ujian, C_i adalah anggaran bilangan orang dalam gambar *i*th, dan C_i^{GT} adalah bilangan sebenar orang (GT) dalam gambar *i*th. Secara umum, MAE menunjukkan ketepatan anggaran manakala MSE menunjukkan ketahanan anggaran. Semakin rendah MAE dan MSE, semakin baik ketepatannya.

$$MAE = \frac{1}{NT}\sum_{i=1}^{NT} | C_i - C_i^{GT} | \quad \ldots(6)$$

$$MSE = \sqrt{\frac{1}{NT}\sum_{i=1}^{NT} | C_i - C_i^{GT} |^2} \quad \ldots(7)$$

5.3 Hasil quantitatif

Kami membandingkan prestasi model cadangan iaitu APDNet dengan kajian yang sebelumnya terhadap set data ShanghaiTech Bahagian A dan Bahagian B. Kaedah sebelum yang dipilih ialah LBP + RR *(Local Binary Pattern and Ridge Regression)*, Zhang et al. (C. Zhang, H. Li, X. Wang, X.Yang 2015), MCNN (Y. Zhang et al. 2016), Switch-CNN (D. B. Sam, S. Surya, R. V. Babu 2017), CSRNet (Y. Li, X. Zhang & D. Chen 2018), DRSAN (L. Liu, H. Wang, G. L. , Wanli Ouyang, L. Lin 2018), PCCNet (J. Gao, Q. Wang, X. Li 2019), PACNN (M. Shi, Z. Yang, C. Xu, Q. Chen 2019) dengan model yang disebut di Bahagian 5.2 iaitu model pengujian dan model cadangan, APDNet. Model perbandingan penanda aras tersebut dipilih terdiri daripada tahun yang berbeza, mula daripada tahun 2012 sehingga tahun 2019. Model kami hanya menjalankan latihan sebanyak 100 epoch disebabkan sumber pengkomputeraan yang terhad.

Dengan hasil yang ada, didapati bahawa APDNet mencapai MAE dan MSE yang terendah di Bahagian A dan yang kedua terendah di Bahagian B set data ShanghaiTech. Dalam perbandingan tersebut, *baseline* mencapai prestasi yang biasa sahaja, tetapi prestasinya turun apabila ditambah dengan konvolusi pendilatan piramid. Hal ini demikian kerana konvolusi pendilatan piramid terlalu fokus kepada skala objek dalam imej dan telah mengabaikan kekacauan objek menyebabkannya gagal membezakan latar depan dan latar belakang. Bagi mengatasi masalah tersebut, untuk membezakan latar belakang dan latar depan, dan akhirnya ditambah dengan rangkaian tumpuan. Hasil menunjukkan rangkaian tumpuan dapat membantu model untuk mengingkatkan ketepatannya terhadap pengiraan himpunan orang ramai.

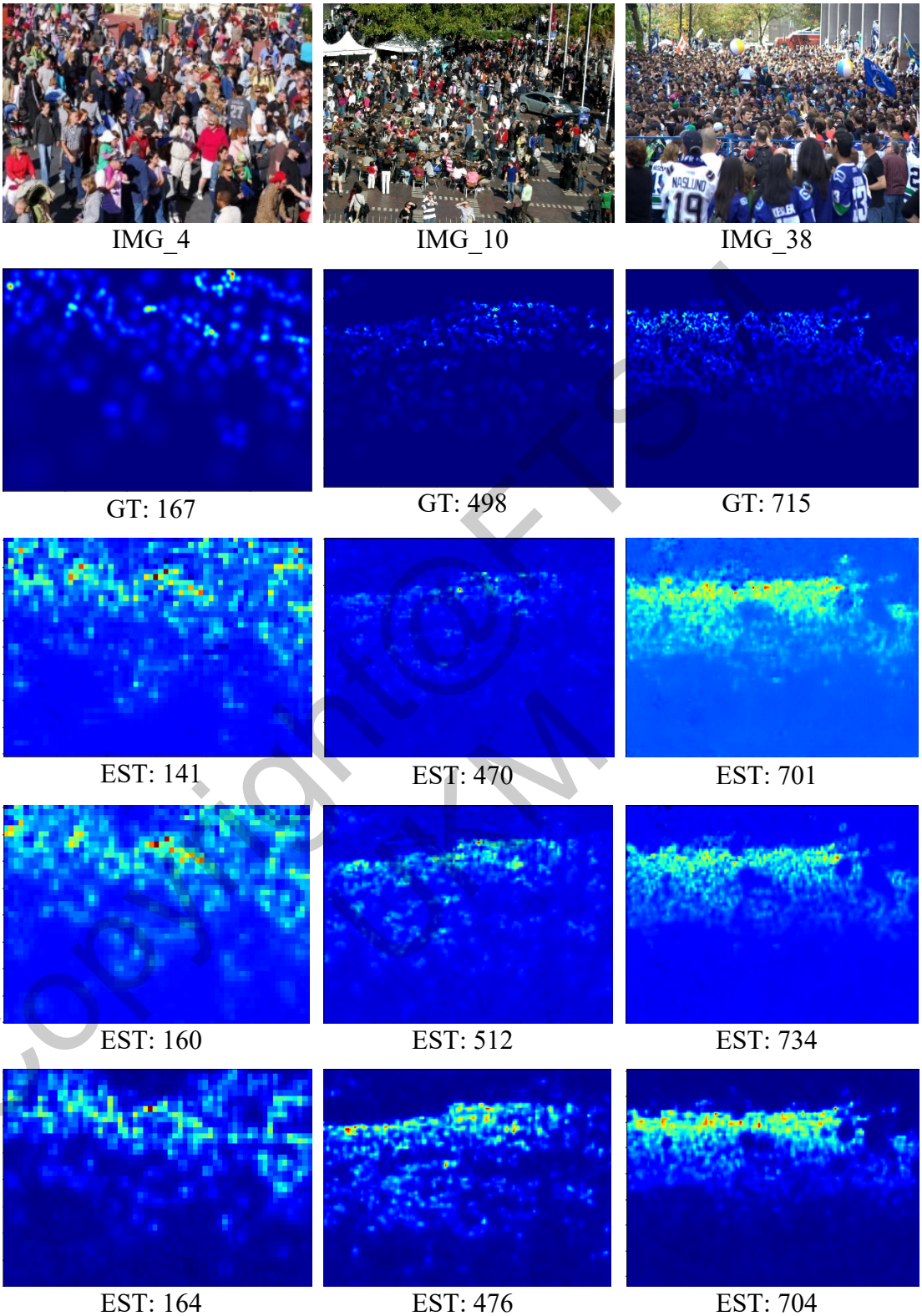
Bukan itu sahaja, projek ini mendapati bahawa rangkaian laju tunggal *(single-column)* mencapai prestasi yang lebih baik berbanding dengan rangkain laju berbilang *(multi-column)*. Hal ini demikian kerana laju tunggal menggunakan lapisan yang lebih kurang daripada laju berbilang. Hasil quantitatif akan dipersembahkan di Jadual 2 dengan perbandingan antara beberapa kaedah.

Kaedah	Seni Bina Rangkaian	Bahagian A		Bahagian B	
		MAE	MSE	MAE	MSE
LBP + RR (2012)	Asas	303.2	371.0	59.1	81.7
Zhang et al. (2015)	Asas	181.8	277.7	32.0	49.8
MCNN (2016)	Lajur berbilang	110.2	173.2	26.4	41.3
Switch-CNN (2017)	Lajur berbilang	90.4	135.0	21.6	33.4
CSRNet (2018)	Lajur tunggal	68.2	115.0	10.6	16.0
DRSAN (2018)	Lajur berbilang	69.3	**96.4**	11.1	18.2
PCCNet (2019)	Lajur berbilang	73.5	124.0	11.0	19.0
PACNN (2019)	Lajur tunggal	66.3	106.4	**8.9**	**13.5**
Baseline	Lajur tunggal	74.6	115.2	12.51	18.41
Baseline + Konvolusi pendilatan piramid	Lajur tunggal	75.6	118.9	16.06	25.13
Baseline + Rangkaian tumpuan + Konvolusi pendilatan piramid **(APDNet - Model cadangan)**	Lajur tunggal	**64.8**	105.6	9.83	15.20

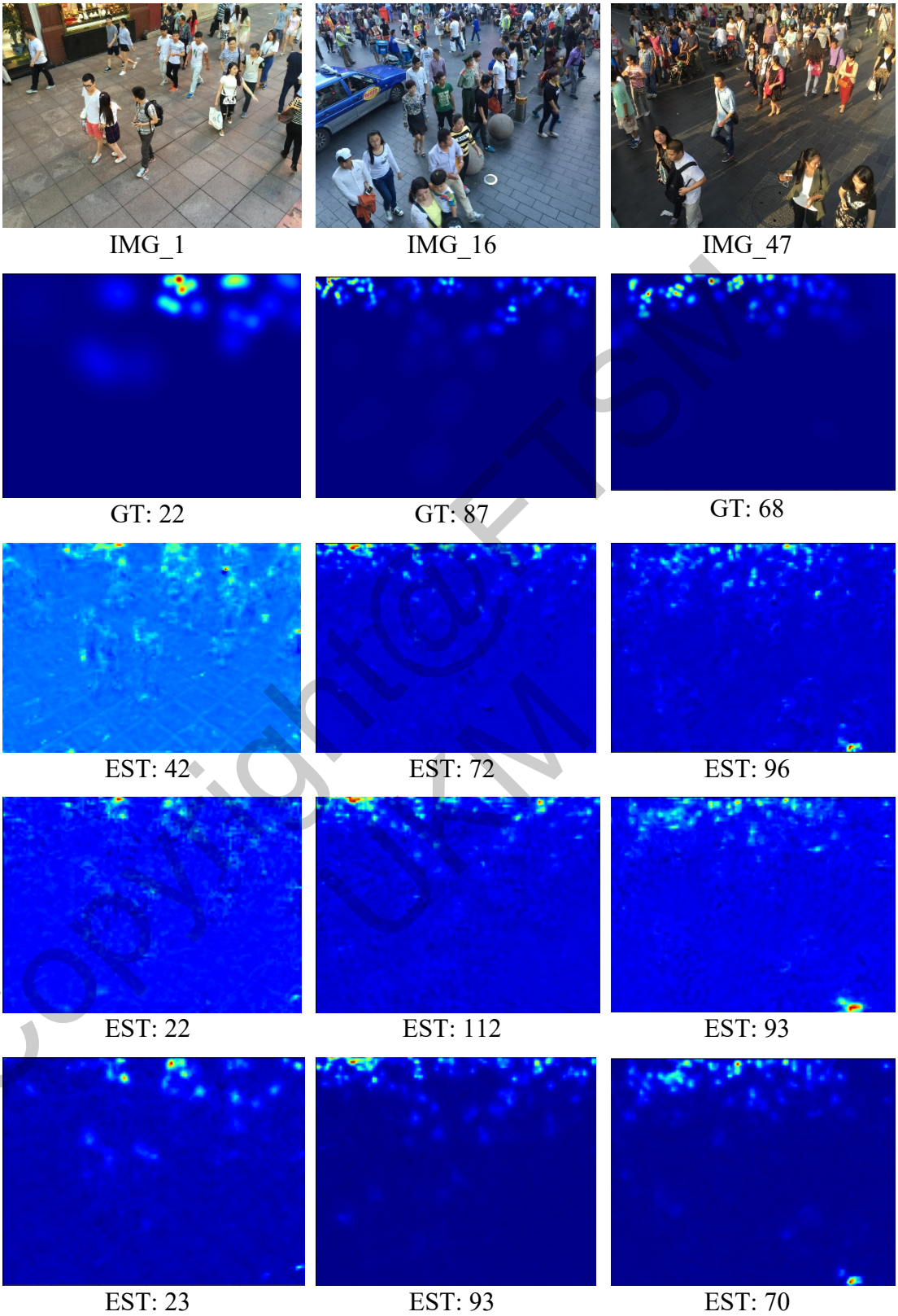
Jadual 2: Perbandingan prestasi pada pengiraan himpunan orang ramai set data ShanghaiTech Bahagian A dan Bahagian B. Prestasi terbaik diserlahkan. Semakin rendah nilai MAE dan MSE, semakin baik model.

5.3 Hasil quanlitatif

Projek diteruskan dengan menjalan analisis kualitatif untuk menyelidiki prestasi APDNet iaitu model cadangan. Beberapa perbandingan kualitatif antara rangkaian *Baseline*, *Baseline* + Konvolusi pendilatan piramid dan APDNet kami dibentangkan di Rajah 2. Perbezaan antara model tersebut boleh merujuk Jadual 1. Di samping itu, beberapa imej dari set data ujian telah dipilih untuk menjalankan ujian qualitatif tersebut. Hasil tersebut akan memaparkan kebenaran dasar dan anggaran himpunan orang ramai di dalam imej. Dari visualisasi Rajah 7, APDNet kami lebih kuat dari segi skala objek, maklumat skala tersebut dicatatkan dengan segi empat merah tersebut. Terdapat bukti yang jelas iaitu dari segi empat merah pertama menunjukkan model kami dapat menangkap skala yang besar. Selain itu, model kami juga dapat menangkap skala yang kecil dengan merujuk kepada segi empat merah di lajur ketiga (IMG_38). Oleh itu, APDNet adalah model yang skala tetap *(scale-invariant)* untuk anggaran kepadatan himpunan orang ramai.



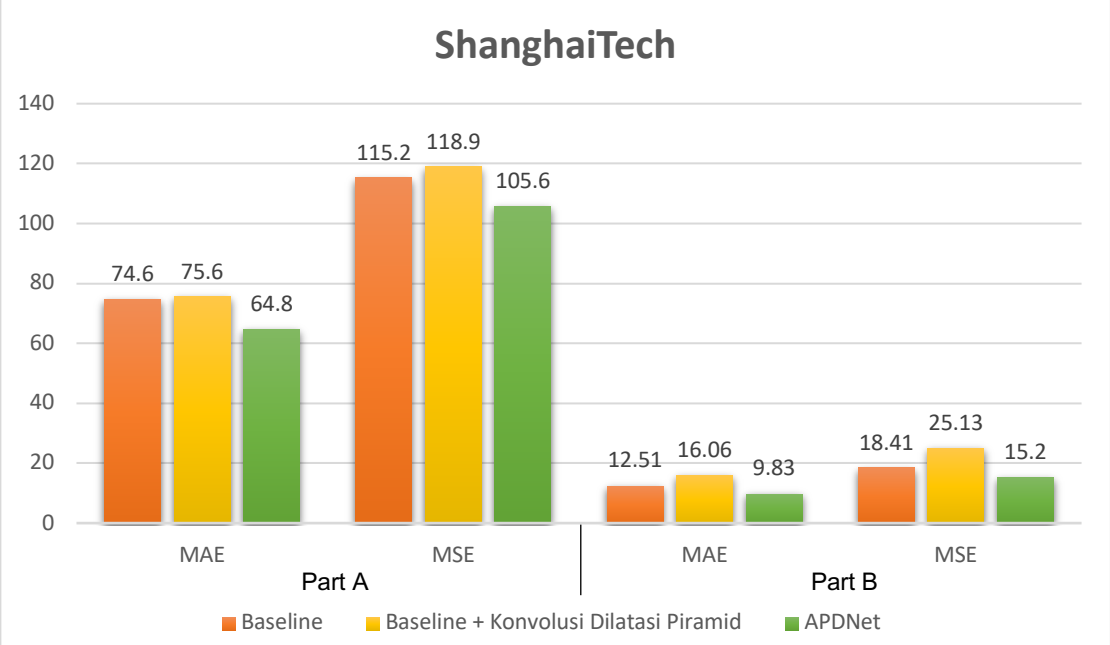
Rajah 7: Visualisasi peta ketumpatan bagi ShanghaiTech Bahagian A. Baris pertama menunjukkan gambar asal. Peta ketumpatan di baris kedua adalah kebenaran dasar, baris ketiga ialah *baseline*, baris keempat ialah *baseline* + konvolusi pendilatan piramid dan baris tera terakhir ialah APDNet kami. "GT" dan "EST" masing-masing menunjukkan jumlah kebenaran dasar dan jumlah anggaran.



Rajah 8: Visualisasi peta ketumpatan bagi ShanghaiTech Bahagian B. Baris pertama menunjukkan gambar asal. Peta ketumpatan di baris kedua adalah kebenaran dasar, baris ketiga ialah *baseline*, baris keempat ialah *baseline* + konvolusi pendilatan piramid dan baris tera terakhir ialah APDNet kami. "GT" dan "EST" masing-masing menunjukkan jumlah kebenaran dasar dan jumlah anggaran.

5.3 Analisis model

Bahagian ini akan menganalisis kesan setiap komponen terhadap model yang dicadangkan. Model pengujian akan dibincang berdasarkan prestasi masing-masing, iaitu model pertama yang terdiri daripada rangkaian VGG-16 yang telah diubahsuai seperti dijelaskan di Bahagian 4.3.1. dan dilabel sebagai "*baseline*". Seterusnya, model kedua iaitu "*Baseline* + Konvolusi pendilatan piramid" merupakan konvolusi pendilatan piramid ditambah ke *Baseline* tersebut dan akhirnya "*Baseline* + Rangkaian tumpuan + Konvolusi dilatsi piramid" adalah model cadangan atau dinamakan "Attention Pyramid Dilated Network" (APDNet) dengan gabungan *Baseline*, rangkaian tumpuan dan konvolusi pendilatan piramid. Rajah 9 menunjukkan prestasi model pengujian.



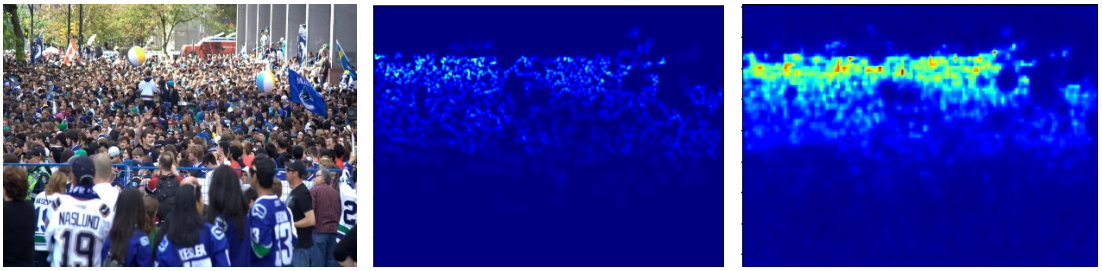
Rajah 9: Prestasi model ujian dengan model cadangan.

I. Kesan baseline: Terdiri daripda tulang belakang VGG-16 tanpa menggunakan lapisan "fully-connected". Model ini mencapai MAE dengan 74.6 sahaja dalam ujian Bahagian A dan 12.51 MAE dalam ujian Bahagian B. Model ini adalah tidak cukup baik kerana ia hanya boleh mengekstrak ciri-ciri asas imej seperti garisan, titik, lengkung, gumpalan

dan sudut. Oleh itu, peta ciri hasilnya juga kabur seperti baris ketiga Rajah 7.

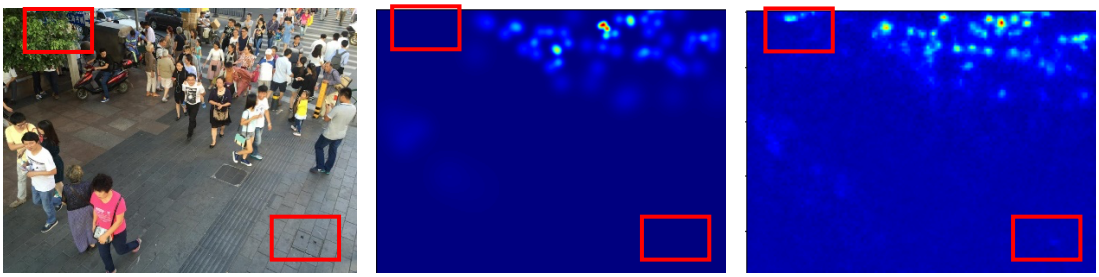
II. Kesan *Baseline* + Konvolusi pendilatan piramid: Model ini diterapkan dengan konvolusi pendilatan untuk meningkatkan medan penerimaan *(receptive field)* dan saiz kernel yang berbeza untuk mempelajari saiz skala yang berbeza. Antara kadar pendilatan yang digunakan ialah 2, 4, 8 dan 12 dan empat lapisan tesebut adalah selari oleh membetuk satu piramid. Kesan konvolusi pendilatan piramid menurunkan prestasi baseline dengan mencapai 75.6 MAE dalam ujian Bahagian A dan 16.06 MAE dalam ujian Bahagian B, oleh kerana model ini tidak dapat membezakan latar depan dan belakang imej dan hanya mengekstrak ciri skala besar dan kecil. Tetapi, dengan tambahan konvolusi pendilatan piramid, model dapat mempelajari ciri skala yang bebeza seperti objek yang lebih dekat kepada camera.

III. Kesan *Baseline* + Rangkaian tumpuan + Konvolusi pendilatan piramid: Model ini menggabungkan model kedua dengan menerapkan rangkaian tumpuan di antara *Baseline* dan konvolusi pendilatan piramid. Hal ini demikian, model ini dipilih sebagai model cadangan, APDNet. Rangkaian tumpuan bertujuan untuk membolehkan model untuk membezakan latar depan dan belakang imej sebelum mengekstrak ciri-ciri teliti. Model ini juga menggunakan lajur tunggal kerana kaedah sebelumnya menunjukkan keputusan lajur tunggal lebih baik daripada lajur berbilang. Kesannya dapat dinampak dengan jelas seperti di Rajah 7. Peta ciri yang dihasilkan oleh APDNet lebih serupa dengan kebenaran dasar dan mencapai prestasi yang baik daripada model sebelumnya. APDNet telah mencapai 64.8 MAE dalam ujian Bahagian A dan 12.51 MAE dalam ujian Bahagian B. Ia bukan sahaja dapat mengekstrak ciri jauh tetapi juga ciri dekat.



IMG_38 GT: 715 EST: 704

Rajah 10: Visualisasi peta ketumpatan bagi Image_38 dalam set data ujian daripada model. "GT" dan "EST" masing-masing menunjukkan jumlah kebenaran dasar dan jumlah anggaran.



IMG_85 GT: 58 EST: 48

Rajah 11: Visualisasi peta ketumpatan bagi Image_85 dalam set data ujian daripada model. "GT" dan "EST" masing-masing menunjukkan jumlah kebenaran dasar dan jumlah anggaran.

Image_38 terdiri daripada objek manusia yang pelbagai skala seperti orang yang dekat kepada camera menunjukkan skala yang besar manakala orang yang jauh kepada camera menunjukkan skala yang kecil. Rajah 10 menunjukkan model cadangan, APDNet dapat mengekstrak ciri-ciri dengan teliti walaupun kepadatan himpunan orang ramai tersebut adalah sangat padat. Namun begitu, terdapat kelemahan dalam model ini iaitu model akan menganggap objek yang serupa dengan kepala manusia akan dikira juga seperti yang ditunjukkan di Rajah 11. Di Rajah 11, model salah meramal pokok tersebut sebagai manusia kerana mereka mempunyai bentuk yang sama dan hujung yang serupa tinggi seperti dalam kotak merah walaupun tiada manusia dalam kebenaran dasar. Di samping itu, terdapat titik di bawah peta ciri (bawah kanan), walaupun tiada manusia sebenarnya. Ini membuktikan model menunjukkan ramalan yang rendah di kepadatan manusia yang jarang kerana set data yang dikumpul kebanyakkan terdiri daripada manusia yang padat dan model salah meramal titik lantai tersebut sebagai manusia. Ramalan yang salah telah mempengaruhi ketepatan model terhadap pengiraan himpunan orang ramai.

6 KESIMPULAN

Secara kesimpulan, projek ini bertujuan untuk menghasilkan satu seni bina model pembelajaran mendalam yang mampu untuk mengatasi variasi skala dalam gambar dalam tugasan pengiraan himpunan orang ramai. Beberapa pendekatan yang sedia ada menunjukkan bahawa analisis orang ramai perlu dikaji mendalam kerana pendekatan yang sedia ada masih mempunyai kelemahan ataupun ruangan yang boleh ditingkatkan. Oleh itu, seni bina model pembelajaran mendalm iaitu model cadangan, APDNet telah dicadangkan supaya dapat meningkatkan keupayaan model dengan memberi tumpuan untuk membezakan latar depan dan belakang dalam gambar dan seterusnya mengatasi skala yang berbeza dalam gambar. Keseluruhan projek telah dibangunkan dalam tempoh masa yang ditetapkan dan dapat mencapai objektif kajian.

Namun begitu, projek ini mempunyai beberapa batasan yang dihadapi dan cadangan untuk menambahbaikkan akan dibincangkan. Batasan pertama dalam projek ini adalah sumber pengkomputeran. Model cadangan memerlukan gpu yang tinggi dilatih oleh kerana gpu yang rendah digunakan menyebabkan ia sangat memakan masa dan model dijangka dilatih dalam masa 3 hari berikutan kerana terlalu banyak penapis dan kerumitan model cadangan.

Di samping itu, model ini boleh ditambahbaik dengan menambahkan bilangan set latihan. Tambahan pula, oleh kerana batasan gpu yang ada, model ini hanya dilatih sampai 100 epoch. Keperluan perkakasan perlu ditingkatkan dengan menggunakan komputer makmal yang mempunyai perkakasan yang tinggi. Di samping itu, untuk mempercepatkan proses latihan, pelantar dan persekitaran pengaturcaraan perlu ditukar kepada pengkomputeran awan yang lebih profesional.

7 RUJUKAN

Raissa Kasolowsky. 2015. Death toll in Saudi haj disaster at least 2,070: Reuterstally Reuters. *https://cn.reuters.com/article/cnews-us-saudi-haj-idCAKCN0SN2F020151029* [30 September 2020].

Tran Phuoc Bao Thu, Pham Nguyen Hong Ngoc, Nhuyen Minh Hai and LeAnh Tuan. 2020. Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries. *Science of The Total Environment, Volume 742, pp. 140430.*

Yuee Huang, Tan Xu and Wen-jie Sun. 2015. Public Health Lesson from ShangHai New Year's Eve Stampede. *Journal of Iran J Public Health. Vol. 44, No.7 pp.1021-1022.*

Payman Salamati, Vafa Rahimi-Movaghar. 2016. Need for Revision of Strategies for Prevention of Hajj Stampedes. *Archives of Trauma Research. In Press. 10.5812/atr.36308.*

Chong Shang, Haizhou Ai, Bo Bai. 2016. End-to-end Crowd Counting Via Jiont Learning Local and Global Count. *In IEEE ICIP, pages 1215-1219.*

Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, Yi Ma. 2016. Single-Image Crowd Counting Via Multi-Column Convoluntional Neural Network. *In Proceedings of the IEEE CVPR pages. 589–597.*

Lingke Zeng, Xiangmin Xu, Bolun Cai, Suo Qiu, Tong Zhang. 2017. Multi-Scale Concolutional Neural Networks For Crowd Counting. *IEEE International Conference on Image Processing (ICIP), 2017, pp. 465-469.*

Tao Zhao, Ram Nevatia and Bo Wu. 2008. Segmentation and Tracking of Multiple Humans in Crowded Environments. *In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 7, pp. 1198-1211.*

Weina Ge, Robert T. Collins. 2009. Marked Point Processes For Crowd Counting. *IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 2913-2920.*

Haroon Idrees, Imran Saleemi, Cody Seibert, Mubarak Shah. 2013. Multi-Source Multi-Scale Counting in Extremely Dense Crowd Images. *IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, 2013, pp. 2547-2554.*

Viet_Quoc Pham, Tatsuo Kozakaya, OsamuYamaguchi, Ryuzo Okada. 2015. COUNT Forest: CO-Voting Uncertain Number of Targets Using Random Forest for Crowd Density Estimation. *IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 3253-3261.*