BAB II
TINJUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Pada tinjauan pustaka ini, penulis mendeskripsikan perkembangan penelitian yang penulis jadikan referensi. Tentang masalah yang diangkat dan yang belum terpecahkan dalam penelitian sebelumnya serta meyakinkan kelayakan penelitian ini. Beberapa penelitian menerapkan *deep learning* dari model *MobileNetsV2* (Sandler dkk., 2018) dan *NasNet* (Zoph dkk., 2018) evaluasi performa dan akurasi untuk mendeteksi ekspresi pada wajah adalah sebagai berikut.

Hal ini telah menginspirasi peningkatan penelitian tentang penggunaan pembelajaran mendalam dalam domain pemrosesan gambar dan visi komputer. Makalah ini menyajikan studi tentang penggunaan pendekatan berbasis *deep learning* untuk mengidentifikasi tanaman yang sakit menggunakan gambar daun dengan transfer *learning*. Studi ini menggunakan arsitektur *NASNet* untuk *Convolutional Neural Network* (CNN). Model tersebut kemudian dilatih dan diuji menggunakan kumpulan data proyek *PlantVillage* yang tersedia untuk umum yang berisi berbagai gambar daun tanaman dengan berbagai variasi status dan lokasi infeksi pada tanaman. Dengan menggunakan model tersebut, tingkat akurasi mencapai 93,82% (Adedoja dkk., 2019).

The Facial Expression Recognition (FER) yang akurat pada ponsel cerdas berguna dalam banyak aplikasi yang merespons keadaan emosi pengguna. Namun, daya komputasi yang lebih terbatas yang tersedia pada perangkat ini membatasi kompleksitas algoritma yang dikembangkan. Penelitian ini memperkenalkan model *Deep Learning* baru yang ringan yaitu *MobiExpressNet*, untuk FER. Model ini mengandalkan konvolusi yang dapat dipisahkan secara mendalam untuk membatasi kompleksitas, dan kami mengadopsi pendekatan *downsampling* yang cepat bersama dengan beberapa lapisan dalam arsitektur untuk menjaga ukuran model tetap sangat kecil. Melalui eksplorasi kemungkinan variasi dalam struktur model, pada penelitian ini, peneliti menentukan bahwa model jaringan terbaik memberikan akurasi 67,96% pada *dataset* FER2013 yang melebihi akurasi manusia sebesar

2,5%. Ukuran model MobiExpressNet dan FLOP ditampilkan lebih dari 5 kali lebih kecil dari model MobileNet terkecil yang membuat model yang dikembangkan sangat menarik untuk aplikasi real-time (Cotter, 2020).

Penelitian ini mengevaluasi kinerja *Neural Architecture Search Network* (*NASNet*) dalam deteksi otomatis COVID-19 (*Coronavirus Disease* 2019) dari gambar rontgen dada. COVID-19 adalah penyakit yang disebabkan oleh *Severe Acute Respiratory Syndrome Coronavirus* 2 (*SARS-CoV-2*) yang menyebabkan pasien mengalami demam, batuk, sesak napas, nyeri otot, produksi dahak, diare, bahkan sakit tenggorokan. Virus ini menyebar melalui udara, dan hingga saat ini meluas sebagai pandemi global. Tidak ada vaksin dan berakibat fatal bagi sekitar 2-7% populasi yang terinfeksi. Di antara karakteristik klinis dan paraklinis pasien yang terinfeksi, nodul telah diidentifikasi dalam gambar rontgen dada yang dapat diidentifikasi secara visual, menghasilkan metode identifikasi yang sederhana, cepat, dan tersedia secara umum. Namun, penyebaran penyakit yang cepat berarti kurangnya tenaga medis khusus yang mampu mengidentifikasinya, itulah sebabnya skema otomatis dikembangkan. Peneliti mengusulkan penyetelan model konvolusional tipe *NASNet* untuk secara otomatis menentukan keadaan awal pasien dalam proses triase atau protokol intervensi pusat perawatan kesehatan. Jaringan saraf dilatih dengan gambar publik dari kasus yang secara positif diidentifikasi sebagai pasien yang terinfeksi virus dan pasien dalam kondisi normal tanpa infeksi. Evaluasi kinerja juga dilakukan dengan citra nyata yang tidak diketahui model neuronalnya. Sedangkan untuk metrik kinerja, kami menggunakan fungsi *loss of cross-entropy* (kategorikal cross-*entropy*), akurasi (atau tingkat keberhasilan), dan MSE (*Mean Squared Error*). Model yang disetel mampu mengklasifikasikan gambar uji dengan benar dengan akurasi 97% (Martínez, dkk., 2020).

Bahasa isyarat adalah bentuk bahasa komunikasi yang dirancang untuk menghubungkan penyandang tuna rungu-tuna rungu dengan dunia. Untuk mengungkapkan suatu ide diperlukan penggunaan gestur tangan dan gerak tubuh. Namun, sebagian besar populasi umum tetap tidak berpendidikan untuk memahami bahasa isyarat. Oleh karena itu, dibutuhkan penerjemah untuk memfasilitasi komunikasi tersebut. Penelitain ini ini ingin memperluas modeldiusulkan

sebelumnya *Convolutional Neural Network* (CNN) yanguntuk memprediksi Bahasa Isyarat Amerika dengan model pembelajaran transfer berbasis *MobileNetV2*. Model terakhir secara efektif digeneralisasikan pada dataset yang berukuran sekitar 18 kali lebih besar dengan 5 kelompok tambahan tanda tangan. Lebih dari 98% akurasi pengenalan telah dilaporkan. Karena parameternya yang relatif lebih sedikit dan operasi komputasi yang kurang intensif dibandingkan dengan arsitektur pembelajaran dalam lainnya, model ini juga ideal untuk diterapkan pada perangkat seluler. Model ini akan berfungsi sebagai kunci untuk menerapkan perangkat lunak penerjemah bahasa isyarat pada ponsel cerdas untuk meningkatkan efisiensi komunikasi antara penyandang tuna rungu dan masyarakat umum (Lum dkk., 2020).

2.2. Dasar Teori

Pada tahapan ini, penulis memaparkan dasar teori yang berkaitan dengan peneltian yang dikembangkan dan model yang digunakan pada penelitian evaluasi performa dan akurasi dari model *MobileNetsV2* dan *NASNet* untuk mendeteksi emosi melalui citra wajah.

2.2.1. *Deep Learning*

Deep learning, kelas teknik *machine learning* yang digunakan untuk mengekstrak fitur dari data (Nishani & Cico, 2017). *Deep learning* merupakan kelas teknik *machine learning*, di mana banyak lapisan tahapan pemrosesan informasi dalam arsitektur hierarkis dimanfaatkan untuk pembelajaran fitur tanpa pengawasan dan untuk analisis/klasifikasi pola. Inti dari pembelajaran mendalam adalah untuk menghitung fitur atau representasi hierarki dari data pengamatan, di mana fitur atau faktor tingkat yang lebih tinggi ditentukan dari yang tingkat yang lebih rendah (Deng, 2014). Komputer dilatih untuk menggunakan kumpulan data besar dan kemudian mengubah nilai piksel gambar menjadi representasi internal di mana pengklasifikasi dapat mendeteksi pola pada input (Azizah dkk., 2017).

Deep learning adalah pembelajaran representasi multi-layer dalam jaringan saraf tiruan (LeCun dkk., 2015). Sedangkan representation learning itu sendiri adalah metode dalam machine learning untuk secara otomatis

mengekstrak/mempelajari representasi (fitur) dari data mentah. Representasi data mentah kemudian dapat digunakan untuk tugas pengenalan atau klasifikasi. Beberapa arsitektur pembelajaran mendalam yang mendasar untuk instance adalah *Convolutional Neural Network* (CNN), *Deep Belief Network* (DBN), *Autoencoder* (AE), dan *Recurrent Neural Network* (RNN). Meskipun merupakan ide lama, Tiga alasan penting untuk popularitas *deep learning* saat ini: pertama, penemuan teknik baru (misalnya, *pretraining* dan *dropout*) dan fungsi aktivasi baru (misalnya ReLU), kedua, pasokan data yang sangat besar (data besar), dan ketiga kemampuan pemrosesan chip yang meningkat secara drastis (misalnya, unit GPU) (Gultom dkk., 2018).

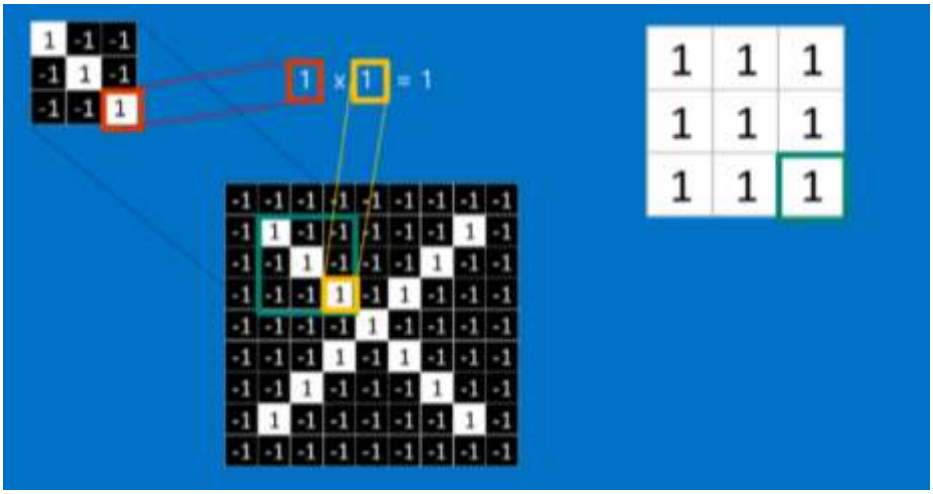
2.2.2. *Convolutional Neural Network*

Convolutional Neural Network (CNN) termasuk dalam jenis deep learning karena kedalaman jaringannya. Deep learning adalah cabang dari machine learning yang dapat mengajarkan komputer untuk melakukan pekerjaan selayaknya manusia, seperti komputer dapat belajar dari proses *training* (Ravi dkk., 2020) CNN dapat dibilang sangat sukses dalam video dan image recognition berskala besar (Pigou dkk., 2015), penerapan CNN telah banyak digunakan oleh industri seperti Amazon, Facebook, dan Google bahkan menggunakan CNN untuk mengambil nomor rumah dari gambar *StreetView* (Lum dkk., 2020). Pada CNN setiap neuron dipresentasikan dalam bentuk 2 dimensi, sehingga metode ini cocok untuk pemrosesan dengan input berupa citra (Maggiori dkk., 2017).

Berbeda dengan manusia, komputer mengenali gambar dalam bentuk array dari nilai piksel-pikselnya. Bayangkan untuk kasus input gambar dengan resolusi pixel 260×260, akan terdapat 260×260×3 array dari angka. Angka 3 yang merupakan suku ketiga dari perkalian melambangkan 3 nilai RGB pada gambar. Selanjutnya, array angka-angka tersebut akan diproses oleh CNN untuk dijadikan nilai kemungkinan suatu gambar berada pada kelas tertentu, misalnya 96% untuk jenis Poodle, 8% untuk jenis Spaniel, dan 0.3% untuk Pomeranian.

CNN terdiri dari dua tahapan utama yaitu *feature learning* dan classification. Pada tahapan *feature learning* terdiri dari *convolution layer*, ReLU (fungsi aktivasi)

dan *pooling layer* sedangkan pada tahap classification terdiri dari *flatten*, *fully-connected layer*, dan prediksi. Pada setiap bagian CNN terdapat dua proses utama, yaitu feed-forward dan backpropagation (Yang dkk., 2016). Typical model CNN ditunjukan gambar 2.1.



Gambar 2.1 *Typical model CNN*

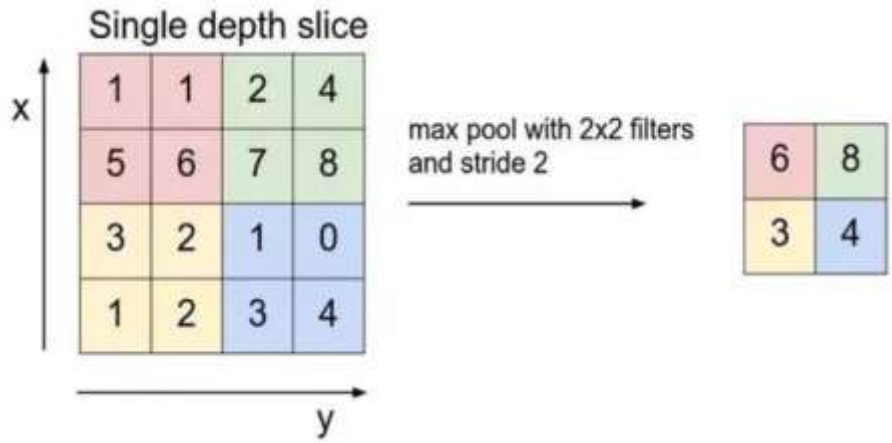
CNN adalah salah satu model *Deep learning*, yang dapat dilihat sebagai ekstraktor otomatis. Ekstraktor fitur berisi lapisan peta fitur dan mengambil fitur pembeda dari gambar mentah melalui tiga lapisan utama: *convolution layers*, *pooling layers*, dan *the fully connected layer*. Operasi khusus sebagai berikut (Yang dkk., 2016):

a. *Convolution Layers*: dilakukan operasi konvolusi yang merupakan proses utama pada CNN. Konvolusi adalah istilah matematis yang berarti mengaplikasikan sebuah fungsi pada output fungsi lain secara berulang [10]. Ketika menguji keberadaan fitur pada gambar baru, CNN akan mencoba semua kemungkinan posisi pada gambar. Filter dibuat untuk menghitung kecocokan fitur pada keseluruhan gambar.

   Gambar 2.1 adalah contoh operasi konvolusi yang membaca matriks 3×3 dan mengalikan matriks tersebut secara *element-wise* dengan dirinya sendiri.

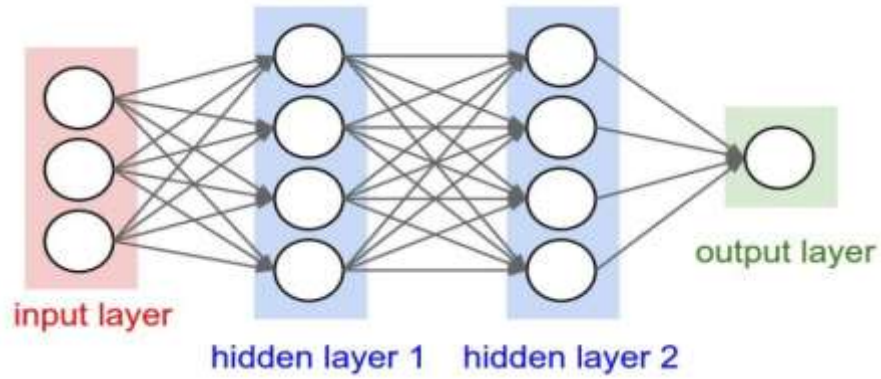
b. *Pooling layers*: dilakukan operasi subsampling, yaitu proses mereduksi ukuran sebuah data citra (Cotter, 2020). Pooling bertujuan untuk memperkecil gambar berukuran besar tetapi masih menyimpan informasi-

informasi pentingnya. Salah satu pendekatan yang banyak digunakan untuk CNN adalah *max pooling*, *max pooling* membagi matriks gambar ke dalam beberapa bagian kecil dan memilih nilai paling besar di dalamnya untuk digunakan ketika matriks gambar baru yang sudah direduksi dibentuk. Gambar 2.2 menunjukkan konsep max pooling dengan filter 2×2 dan pergeseran (*stride*).



Gambar 2.2 Operasi *Max Pooling*

c. *The fully connected layer*: adalah layer yang biasa digunakan pada *Multilayer Perceptron* (MLP) dan bertujuan untuk melakukan transformasi pada dimensi data agar dapat diklasifikasi secara linear. Gambar 2.3 menggambarkan suatu MLP dengan 2 hidden layer yang *fully-connected*. Output dari layer ini adalah array dengan panjang jumlah kelas yang model harus pilih. Pada *fully connected layer*, bobot yang paling besar dari layer sebelumnya akan menentukan fitur mana yang paling berhubungan dengan kelas atau label tersebut.



Gambar 2.3 *Multi-layer Perceptron* Sederhana 2 *Hidden Layer*

2.2.3. Pemodelan *MobileNetV2*

Keberhasilan metode *Deep learning* pada berbagai tugas visi komputer telah menyebabkan penerapannya yang luas (Cotter, 2020) dan gambaran umum penerapannya di *The Face Expression Recognition* (FER) diberikan dalam penelitian (Li dan Deng, 2020). Pada bagian ini, perubahan arsitektur yang diperkenalkan di *MobileNetV1* (Howard dkk., 2017) dan *MobileNetV2* (Sandler dkk., 2018) yang secara signifikan mengurangi biaya komputasi dibandingkan dengan metode *Deep learning* lainnya yang ditujukan untuk pengenalan objek dengan hanya sedikit penurunan kinerja.

MobileNet, merupakan salah satu arsitektur CNN yang dapat digunakan untuk mengatasi kebutuhan akan *computing resource* berlebih. Perbedaan mendasar antara arsitektur MobileNet dan arsitektur CNN pada umumnya adalah penggunaan lapisan atau layer konvolusi dengan ketebalan filter yang sesuai dengan ketebalan dari input image (Howard dkk., 2017).

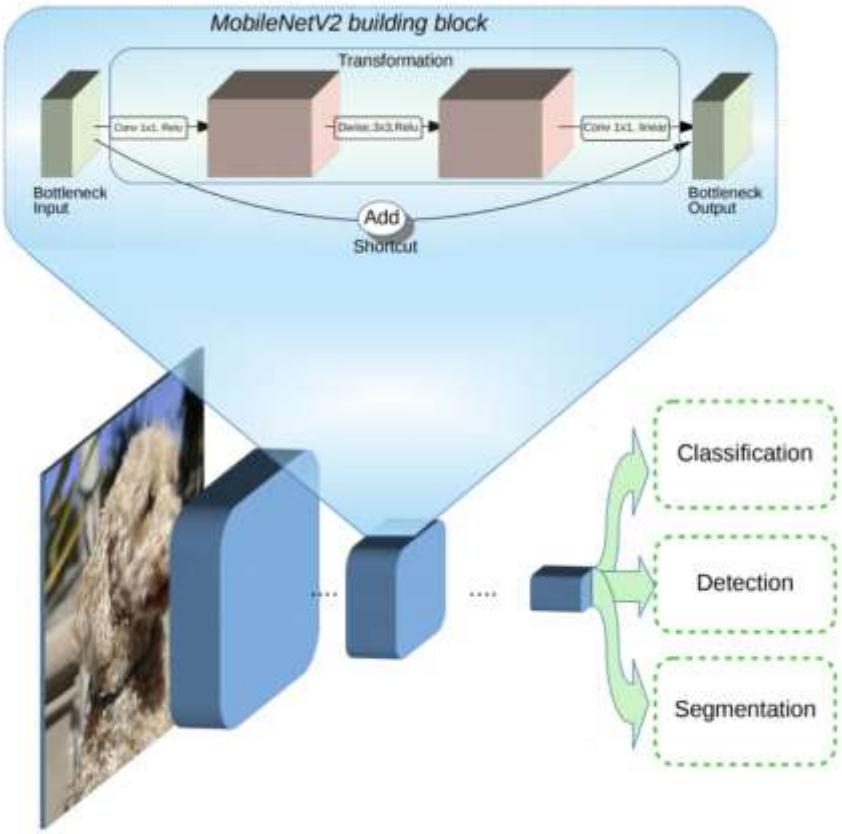
Perbedaan mendasar antara arsitektur MobileNet dan arsitektur CNN pada umumnya adalah penggunaan lapisan atau layer konvolusi dengan ketebalan filter yang sesuai dengan ketebalan dari input image. MobileNet membagi konvolusi menjadi depthwise convolution dan pointwise convolution. Lapisan pertama disebut *depthwise convolution*, karena melakukan penyaringan ringan dengan menerapkan filter konvolusional tunggal per saluran masukan. Lapisan kedua adalah konvolusi 1 × 1, yang disebut *pointwise convolution*, yang bertanggung

jawab untuk membangun fitur baru melalui komputasi kombinasi linier dari saluran input.

Tabel 2.1 Kesulurhan Arsitektur MobilenetV2

No	Input	Operator	t	c	n	s
1	224^2 x 3	conv2d	-	32	1	2
2	112^2 x 32	bottleneck	1	16	1	1
3	112^2 x 16	bottleneck	6	24	2	2
4	56^2 x 24	bottleneck	6	32	3	2
5	28^2 x 32	bottleneck	6	64	4	2
6	14^2 x 64	bottleneck	6	96	3	1
7	14^2 x 96	bottleneck	6	160	3	2
8	7^2 x 160	bottleneck	6	320	1	1
9	7^2 x 320	conv2d 1 x 1	-	1280	1	1
10	7^2 x 1280	avgpool 7 x 7	-	-	1	-
11	1 x 1 x 1280	conv2d 1 x 1	-	k	-	

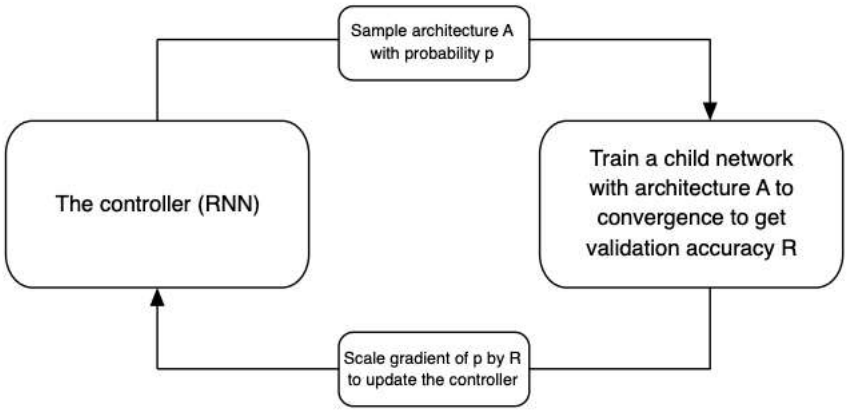
MobileNetV2 meningkatkan kinerja model seluler pada berbagai tugas dan tolok ukur serta di seluruh spektrum ukuran model yang berbeda (Sandler dkk., 2018). Sama seperti *MobileNetV1*, *MobileNetV2* masih menggunakan *depthwise* dan *pointwise convolution*. MobileNetV2 menambahkan dua fitur baru yaitu: 1) *linear bottleneck*, dan 2) shortcut connections antar bottlenecks. Pada bagian bottleneck terdapat input dan output antara model sedangkan lapisan atau layer bagian dalam meng-enkapsulasi kemampuan model untuk mengubah input dari konsep tingkat yang lebih rendah seperti pixels ke deskriptor tingkat yang lebih tingg seperti kategori gambar). Pada akhirnya, seperti halnya koneksi residual pada CNN tradisional, *shortcut* antar *bottlenecks* memungkinkan training atau pelatihan yang lebih cepat dan akurasi yang lebih baik (Sandler dkk., 2018). Pada beberapa penelitian menunjukkan bahwa model *MobiExpressNet*, mencapai akurasi tinggi 67,96% pada set data FER2013 dengan ukuran model sekitar 75000 parameter dan komputasi 1×10^6 FLOP (Cotter, 2020).



Gambar 2.4 Contoh alur MobileNetV2

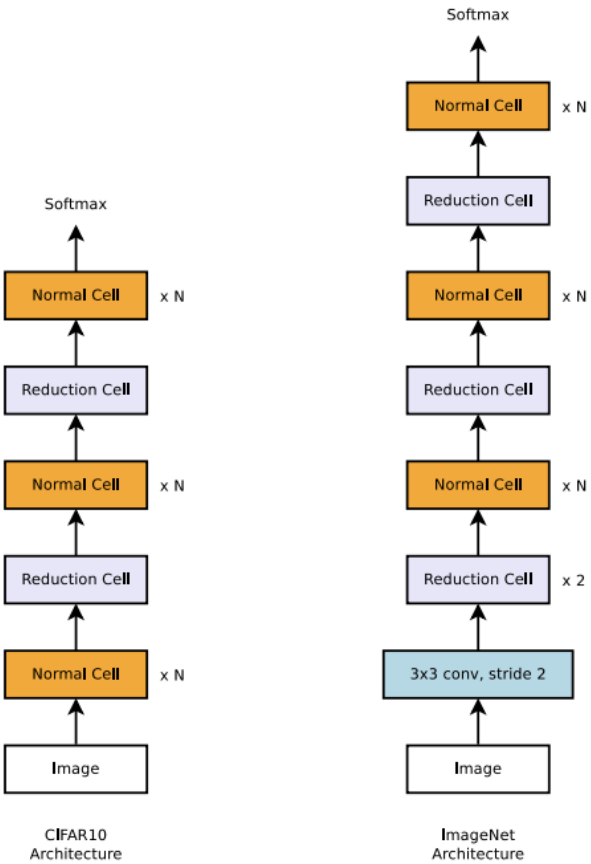
2.2.4. Pemodelan *NasNets*

Neural Architecture Search (NAS) adalah bidang penelitian yang muncul dari berbagai upaya otomatisasi proses desain arsitektur (Kyriakides dan Margaritis, 2020). NAS telah berhasil diterapkan pada arsitektur model desain untuk klasifikasi gambar dan model bahasa (Pham dkk., 2018). NAS telah mampu menghasilkan banyak jaringan canggih, sementara kemajuan di lapangan telah mengusulkan metode yang tidak membutuhkan banyak sumber daya. Ideologi dasar NAS adalah untuk menemukan struktur jaringan kandidat melalui strategi pencarian di ruang pencarian yang ditentukan, berdasarkan umpan balik yang diperoleh dari evaluasi.



Gambar 2.5 penerapan RNN pada Nasnet

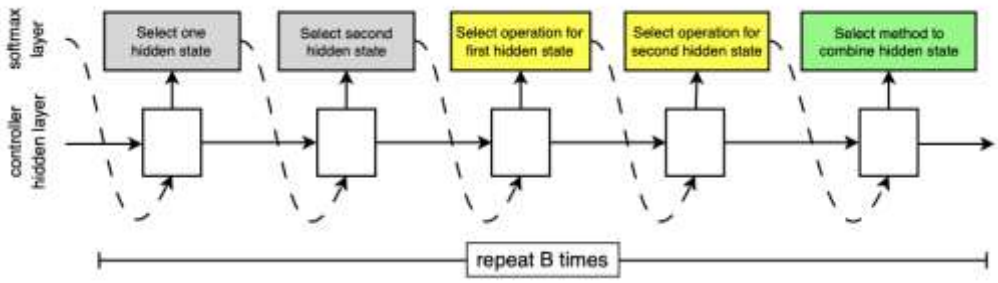
Neural Architecture Search Network (NASNet) sebagai struktur untuk model konvolusional. Jaringan dalam ini diperkenalkan pada awal 2018 oleh tim Google Brain. Dalam desainnya, mereka berusaha untuk menentukan blok bangunan dengan kinerja tinggi dalam kategorisasi sekumpulan gambar kecil (CIFAR-10). Mereka kemudian menggeneralisasi blok ke kumpulan data yang lebih luas (ImageNet) (Martínez, dkk., 2020).



Arsitektur *NASNet* terinspirasi oleh kerangka *Neural Architecture Search* (NAS) (Zoph dan Le, 2017). Arsitektur *deep learning NASNet* sangat fleksibel dan

dapat diskalakan dalam penggunaan sumber daya komputasi dan parameter untuk kasus penggunaan yang berbeda di domain yang berbeda (Adedoja dkk., 2019). NASNet adalah kesadaran bahwa rekayasa arsitektur dengan CNN sering mengidentifikasi motif berulang yang terdiri dari kombinasi bank filter konvolusional, nonlinier, dan pemilihan koneksi yang cermat untuk mencapai hasil yang canggih

Di *NASNets*, RNN mengambil sampel jaringan turunan dengan arsitektur berbeda. Jaringan anak dilatih untuk konvergensi untuk mendapatkan beberapa akurasi pada set validasi yang diadakan. Akurasi yang dihasilkan digunakan untuk memperbarui pengontrol sehingga pengontrol akan menghasilkan arsitektur yang lebih baik dari waktu ke waktu (Zoph dkk., 2018).



Gambar 2.7 RNN saat mencari akurasi tertinggi pada network

2.2.5. *SoftMax Function*

Fungsi Softmax adalah salah satu dari sekian banyak model fungsi aktifasi yang digunakan pada jaringan saraf tiruan, MobileNetv2 dan Nasnet menggunakan softmax untuk menghitung probabilitas dalam klasifikasi hasil perhitungan CNN. Fungsi Softmax manghasilkan output yang merupakan rentang nilai antara 0 dan 1 melalui hubungan (Chigozie dkk., 2018)

$$f(x_i) = \frac{exp\ (x_i)}{\sum_j exp\ (x_j)} \quad -(1.12)$$

Fungsi Softmax digunakan dalam model *multi-class* diaman fungsi tersebtu mengembalikan probabilitas dari setiap class dengan target kelas yang memiliki probabilitas tertinggi.