

Intelligent Integration of Enterprise

Project 2

人類手骨年齡推斷

Age Prediction: Using Human Phalanges of the Hand

Group 5

109034534 梁芷蘋

109034535 葉子匯

109034536 郭芸如

109034551 林奕辰

指導教授：邱銘傳 教授

目錄

1、背景介紹	5
1.前言	5
2.5W1H	5
3.計畫流程	6
二、模型訓練與績效	6
1. 資料蒐集	6
1.1 原始資料集	6
1.2 本計畫使用之訓練集與測試集	6
2. 資料前處理	8
2.1 資料清理	8
2.2 資料標準化	9
2.3 資料擴增	9
2.4 切割訓練、驗證與測試集	11
3. 模型架構	12
3.1 Xception	13
3.2 InceptionV3	15
3.3 InceptionResNetV2	16
4. 超參數調整與模型績效比較	17
4.1 擴增前資料 初步模型超參數調整	17
InceptionResNetV2	17
4.2 擴增後資料 模型超參數調整	20
三、結果與討論	22
四、未來展望	23
五、參考資料	23

圖目錄

圖一、本計畫訓練用人類手骨 X 光片	5
圖二、計畫執行流程圖	6
圖三、訓練集資料與年齡長條圖	7
圖四、無空值資料程式執行結果	8
圖五、無重複值資料程式執行結果	9
圖六、無異常值資料程式執行結果	9
圖七、數據標準化程式運作結果	9
圖八、原訓練集年齡分布(左圖)與增強後訓練集年齡分佈(右圖)	10
圖九、擴增資料圖	10
圖十、模型訓練流程圖	11
圖十一、資料切割程式碼	11
圖十二、資料切割完成後的 DataFrame 形式	11
圖十三、Xception 架構圖	14
圖十四、Extreme Inception 架構圖	15
圖十五、InceptionV3 架構圖	15
圖十六、InceptionResNetV2 架構圖	16
圖十七、擴增前初步模型架構績效	18
圖十八、X1 模型架構圖	19
圖十九、X1 模型訓練過程	19
圖二十、擴增前(左圖)與擴增後(右圖)模型訓練過程比較(均使用 X_1 模型)	20
圖二十一、原訓練集年齡分布(左圖)與增強後訓練集年齡分佈(右圖)	21
圖二十二、擴增後資料模型績效	22
圖二十三、X_a1 模型訓練過程	22

表目錄

表一、5W1H.....	5
表二、原始資料集.....	6
表三、訓練集與驗證集欄位定義與特性.....	7
表四、測試集欄位定義與特性.....	7
表五、資料切割後未擴增前之統計分析表.....	11
表六、資料切割後擴增後之統計分析表.....	12
表七、模型層級與層級說明.....	12
表八、三種模型與 VGG16 比較表.....	16
表九、模型績效比較表.....	17
表十、資料擴增前後測試集預測結果比較.....	20

1、背景介紹

1. 前言

有鑑於目前科技考古以及科技辦案的興起，機器學習於考古學以及刑事案件的應用越來越加興盛。不論是考古現場或是刑事案件現場，手指骨格的年齡判斷相較於其他(如頭骨和骨盆)有更高的難度，培訓人員所花費的時間非常龐大，且就算是經驗豐富的專家也容易因為指骨體積小而造成有誤判或是不客觀的判讀結果出現。針對這些問題，本研究將運用卷積神經網路(CNN, Convolutional Neural Networks)，從 Kaggle 上所能取得的手指骨圖片對模型進行訓練，以達到能協助快速且客觀協助人員進行手指骨骼年齡判斷。



圖一、本計畫訓練用人類手骨 X 光片

2. 5W1H

一般而言，在進行年齡判斷時，由於手骨的表徵不如頭骨與骨盆來的明顯，因此學者專家往往都是在找不到頭骨骨盆或是這兩者已經破損到無法辨識時才會使用手骨進行年齡的判定。而以手骨進行年齡判定時，就算是專家學者也容易有誤判的情況發生，考慮到這些情形，本計畫先以 5W1H 進行問題定義。

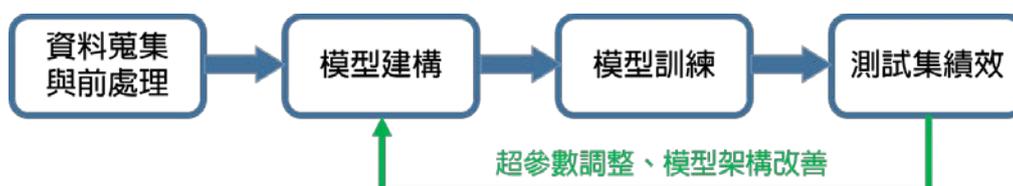
表一、5W1H

Who	刑案現場鑑識人員、法醫、考古學者
What	手骨年齡的推斷
When	1. 發現人類骨骼，警方需要找出與死者相關的身分來推動案件進展 2. 考古墓葬現場整理骨骼，協助建構文化現場的歷史脈絡
Where	可精進刑案現場人員進行骨骼判斷，以及考古遺址現場的骨骸遺留整理
Why	培訓人員所花費的時間非常龐大，且就算是經驗豐富的老手也容易有誤判的情況出現
How	資料分析、深度學習、機器學習

透過電腦輔助進行手骨年齡的判斷，可大幅降低專家學者在進行年齡判斷所需要的培訓時間以及準確度，加速現場分析的進展。

3.計畫流程

本計畫的流程如下圖所示。首先進行相關手骨的 X 光資料收集後，進行資料的前處理。接下來，建構模型並進行訓練，並且測試其績效成果，再根據結果進行超參數調整與模型架構的改善，以達到手骨年齡判定的最佳化。



圖二、計畫執行流程圖

二、模型訓練與績效

1. 資料蒐集

資料來源於 Kaggle 網站的 RSNA Bone Age，資料源自於北美放射學會 Radiological Society of North America (RSNA)的放射訊息委員會 Radiology Informatics Committee (RIC)於 2017 年所開放的資料。

1.1 原始資料集

表二、原始資料集

原始訓練集	boneage-training-dataset (12,611 張手骨 X 光照片)
	boneage-training-dataset.csv (欄位: id、boneage、male)
原始測試集	boneage-test-dataset (200 張手骨 X 光照片)
	boneage-test-dataset.csv (欄位: Case ID、Sex)

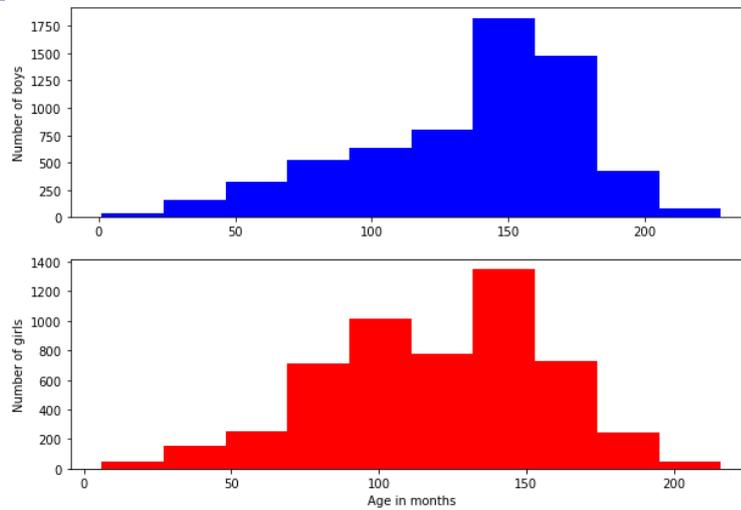
由於原始測試集並未提供「年齡」相關資料，故本研究捨棄原始測試集，從原始訓練集重新切分新的訓練集、測試集。本研究使用之訓練集，測試集資訊詳見資料前處理。

1.2 本計畫使用之訓練集與測試集

- 訓練集與驗證集: 11,611 張手骨 X 光照片

表三、訓練集與驗證集欄位定義與特性

欄位變數	定義	值或特性
id	手骨編號	十進位編號 2491~15610
boneage	受試者實際年齡 (單位:月)	統計資料 count 11611.000000 mean 127.697011 std 40.512637 min 1.000000 25% 96.000000 50% 132.000000 75% 156.000000 max 228.000000
male	是否為男性	TRUE(6285 人) FALSE(5326 人)



圖三、訓練集資料與年齡長條圖

由性別與年齡長條圖可得知，本訓練集資料男性與女性的數量相當，且年齡基本上兩者分布落差不大，是性別分布平均的訓練集資料。

- 測試集: 1,000 張手骨 X 光照片

表四、測試集欄位定義與特性

欄位變數	定義	值或特性
id	手骨編號	十進位編號
boneage	受試者實際年齡 (單位:月)	統計資料 count 1000.000000 mean 122.952000

		std 48.087796 min 4.000000 25% 94.000000 50% 132.000000 75% 162.000000 max 228.000000
male	是否為男性	TRUE(548 人) FALSE(452)

2. 資料前處理

本計畫所使用的資料為 X 光圖片檔，為了日後模型訓練的進展，會先將資料進行前處理，主要分為三個步驟：資料清理、資料標準化與資料切割。

2.1 資料清理

資料清理的主要目的為確認資料沒有下列狀況：

- (1) 錯誤標示的資料 (人工檢查)
- (2) 重複之資料
- (3) 缺失或空的資料值
- (4) 異常值

首先於圖檔資料處理的部分，會先由人工檢查方式確認是否含有全黑，空白，或手部因 X 光圖拍攝不佳不完整的圖片。至於數值資料的部分，則是透過電腦程式進行 csv 資料的前處理，主要透過 Python 的 DataFrame.isnull()與 DataFrame.duplicated()函式，來確認資料集中沒有(3)空值與(2)重複值。接著，由資料分布狀況判斷沒有(4)異常數值如:年齡不為正數等情況出現。

```
In [11]:
...: replicate_list=data_df.duplicated()
...: num_of_replicate=0
...: for i in replicate_list:
...:     if i==True:
...:         num_of_replicate+=1
...: if num_of_replicate==0:
...:     print(f'No replicated data in {len(data_df)} data.')
...: else:
...:     print(f'There are {num_of_replicate} replicated data.')
No replicated data in 12611 data.
```

圖四、無空值資料程式執行結果

```

In [7]: Emptylist=data_df.iloc[:,0:2].isnull()
...:
...: num_of_empty=0
...: for i in Emptylist:
...:     if i==True:
...:         num_of_empty+=1
...: if num_of_empty==0:
...:     print(f'No empty data in {len(data_df)} data.')
...: else:
...:     print(f'There are {num_of_empty} empty data.')
No empty data in 12611 data.

```

圖五、無重複值資料程式執行結果

```

In [13]: data_df.describe()
Out[13]:

```

	id	boneage
count	12611.000000	12611.000000
mean	8537.653001	127.320752
std	4108.763993	41.182021
min	1377.000000	1.000000
25%	5074.500000	96.000000
50%	8565.000000	132.000000
75%	12091.500000	156.000000
max	15610.000000	228.000000

圖六、無異常值資料程式執行結果

2.2 資料標準化

由於數據之間的差異性會影響到後續進行機器學習模型的準確度，因此需要針對這些數據進行標準化的處理。有些數值可能因為其變化範圍較大，而影響到模型預測結果，因此，為了預防這類問題的發生，使用數據標準化來限制權重變化範圍。標準化後的數值，因有較小的變異程度，輸入模型有助於提升模型準確度。

```

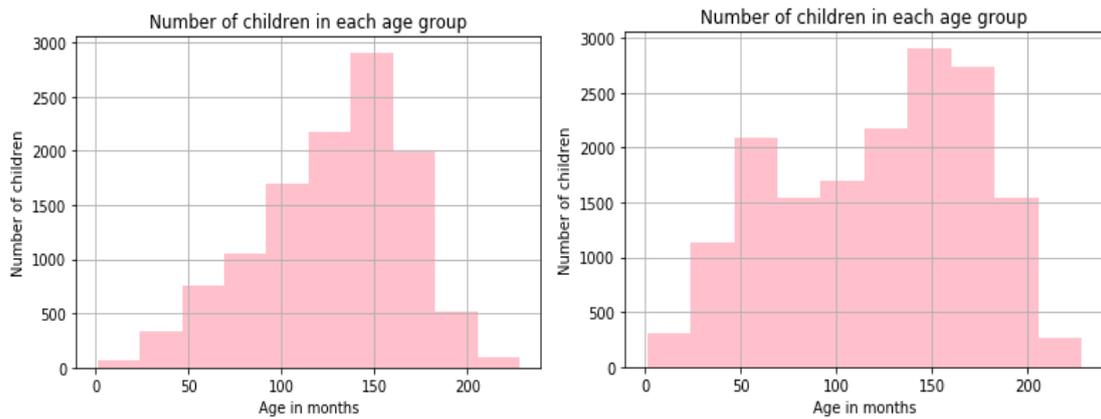
In [18]:
...: std_bone_age = data_df['boneage'].std()
...: data_df['bone_age_z'] = (train_df['boneage'] - mean_bone_age)/(std_bone_age)
...: print(data_df.head())

```

	id	boneage	male	bone_age_z
0	1377	180	False	-1.643849
1	1378	12	False	-0.186902
2	1379	94	False	1.415739
3	1380	120	True	0.541571
4	1381	82	False	0.832960

圖七、數據標準化程式運作結果

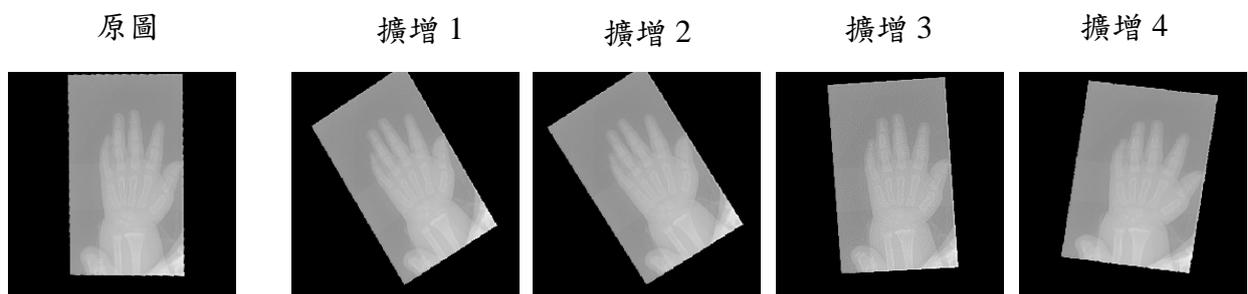
2.3 資料擴增



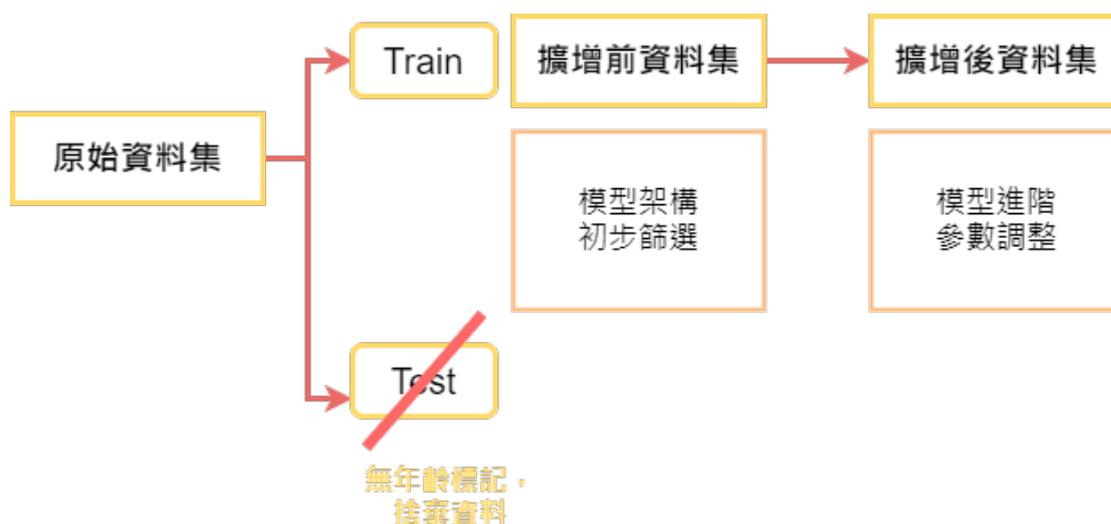
圖八、原訓練集年齡分布(左圖)與增強後訓練集年齡分佈(右圖)

從資料分佈看出在年齡小於 100 個月與大於 175 個月的資料數量明顯較 100-175 個月的少，因此擴增資料集中年齡小於 100 個月與大於 175 個月的資料(透過 `keras.preprocessing.image.ImageDataGenerator()` 函式對圖片旋轉、縮放)，使筆數分佈較平衡，以避免模型在手骨年齡較大或較小時測試績效不佳。

然而，由於擴增前資料筆數龐大共有 11,611 筆，擴增後更高達 16,425 筆，模型訓練時間平均 4~5 小時。為節省模型訓練時間，選擇使用擴增前的資料集先訓練以初步篩選模型結構，再由擴增後的資料集進行訓練，設計出最佳績效的模型。



圖九、擴增資料圖



圖十、模型訓練流程圖

2.4 切割訓練、驗證與測試集

由原始資料共 12,611 筆資料中，隨機取出 1,000 筆作為測試集，再將其餘 11,611 筆資料中，切分 20% 作為驗證集，剩下 80% 為訓練集。

```
df_train,df_valid= train_test_split(data_df, test_size = 1000, random_state = 0)
df_train,df_test= train_test_split(df_train, test_size = 0.2, random_state = 0)
```

圖十一、資料切割程式碼

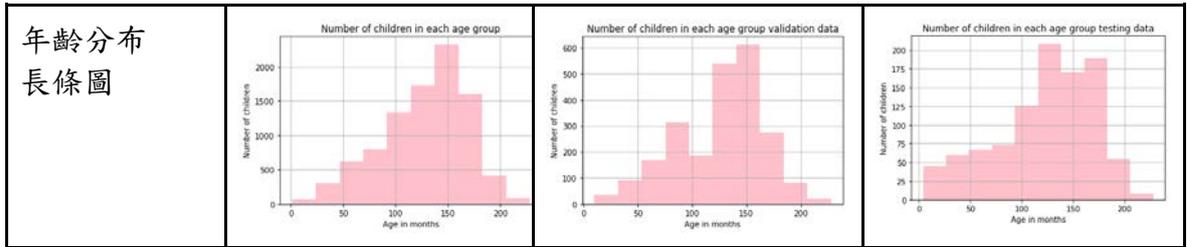
df_test	DataFrame	(1000, 3)	Column names: id, boneage, male
df_train	DataFrame	(9288, 3)	Column names: id, boneage, male
df_valid	DataFrame	(2323, 3)	Column names: id, boneage, male

圖十二、資料切割完成後的 DataFrame 形式

將上述資料進行分析可得到以下表格，由下表資料，各資料集之年齡分佈相近，因此判斷資料已經均勻切割。

表五、資料切割後未擴增前之統計分析表

	訓練集	驗證集	測試集
筆數	9288	2323	1000
年齡統計資料	mean 127.905469 std 40.503867 min 1.000000 25% 96.000000 50% 132.000000 75% 156.000000 max 228.000000	mean 126.863539 std 40.545695 min 10.000000 25% 96.000000 50% 132.000000 75% 156.000000 max 228.000000	mean 122.952000 std 48.087796 min 4.000000 25% 94.000000 50% 132.000000 75% 162.000000 max 228.000000



表六、資料切割後擴增後之統計分析表

	訓練集	驗證集	測試集
筆數	13140	3285	1000
年齡統計資料	mean 122.160198 std 51.959743 min 1.000000 25% 72.000000 50% 132.000000 75% 162.000000 max 228.000000	mean 120.335160 std 52.206596 min 6.000000 25% 72.000000 50% 132.000000 75% 162.000000 max 228.000000	mean 122.952000 std 48.087796 min 4.000000 25% 94.000000 50% 132.000000 75% 162.000000 max 228.000000
年齡分布 長條圖			

3. 模型架構

本研究選定四種在各資料預測競賽中表現優良的 CNN 模型架構，分別是 VGG16、Xception、InceptionV3 與 InceptionResNetV2，來找出能夠得到本資料集最佳預測結果的模型。經嘗試由於 VGG 模型無法在 50 epoch 內收斂，因此本研究主要使用 Xception、InceptionV3 與 InceptionResNetV2 做進一步超參數調整與架構改善，找出績效更佳之模型。

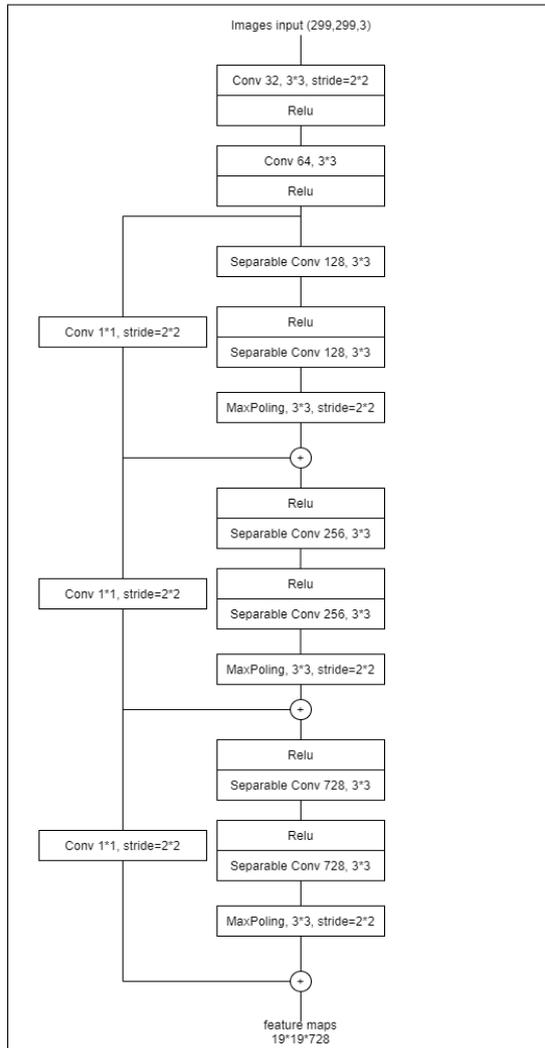
表七、模型層級與層級說明

層級	說明
Keras model	使用 Keras 提供之 CNN 模型(如: Inception、Xception 等)
Pooling	使用 GlobalMaxPooling2D

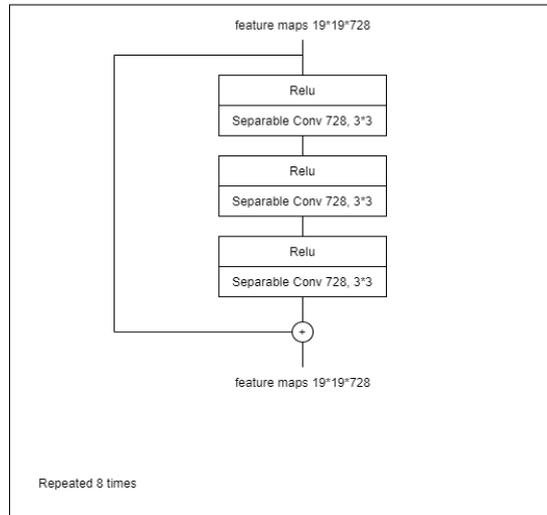
全連接層	Flatten
隱藏層	Dense(10 , activation="relu")
輸出層	Dense(1 , activation="linear")

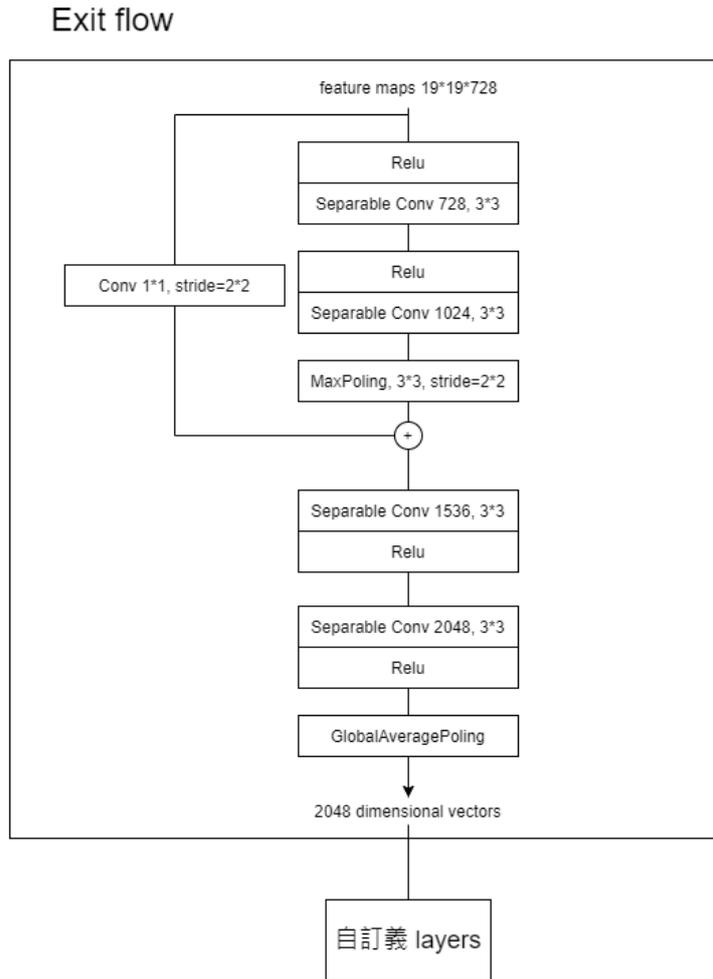
3.1 Xception

Entry flow



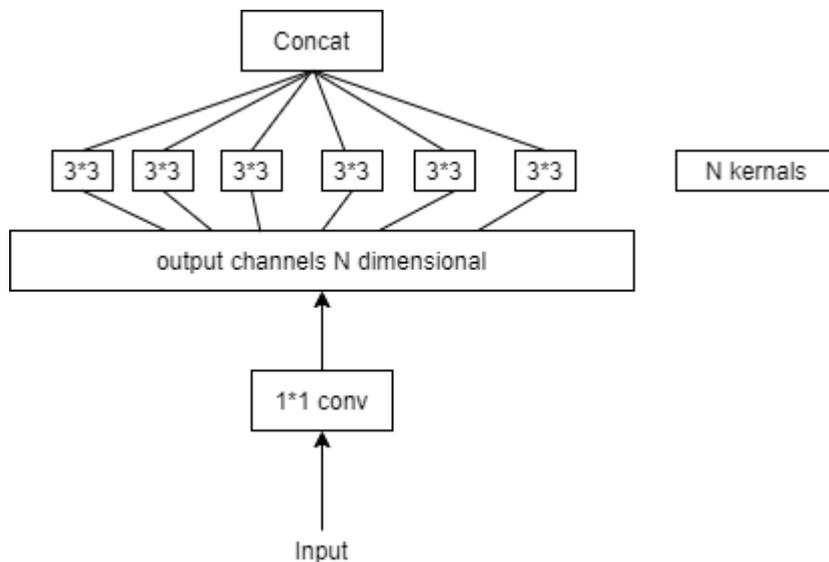
Middle flow





圖十三、Xception 架構圖

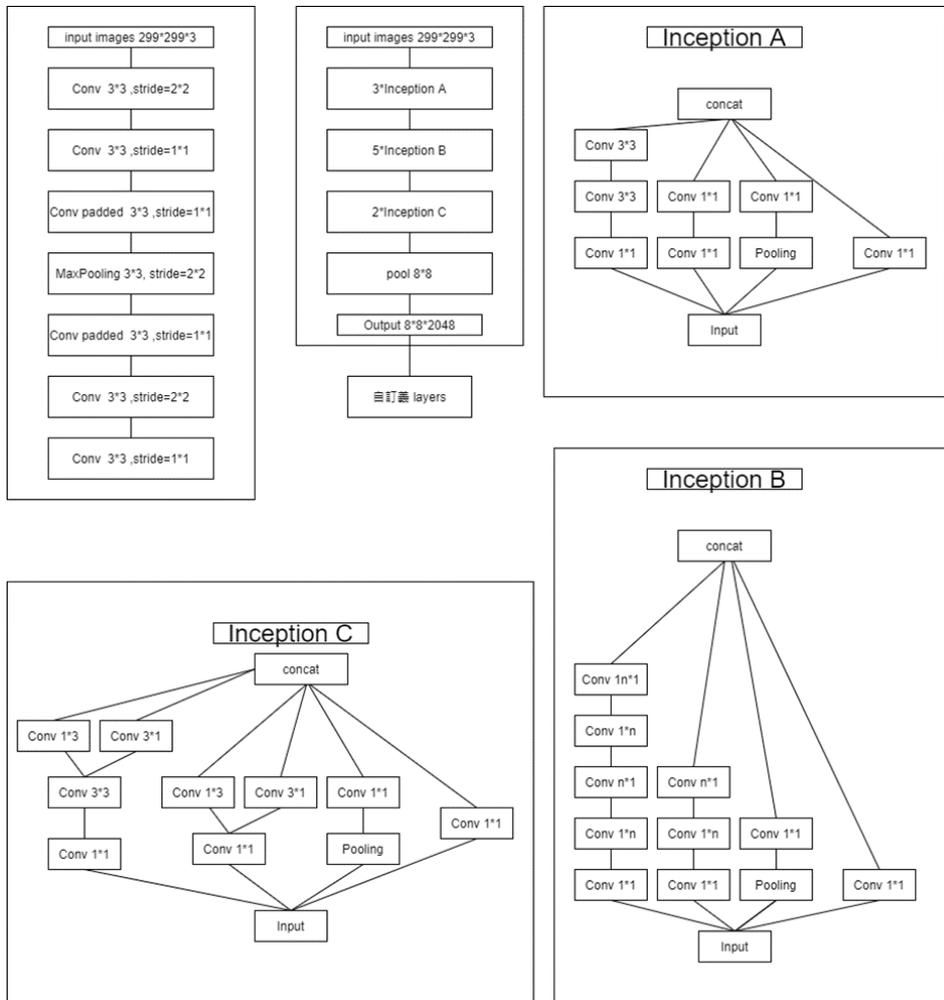
上圖為 Xception net 的基本架構，Xception net 中利用了 depthwise separable convolution (Extreme Inception)的架構，這個計算方式可以幫助該網路藉由較少的計算量得到卷積結果，加速模型的訓練並降低所需要的參數量。



圖十四、Extreme Inception 架構圖

上圖為 Extreme Inception 的示意圖，其核心概念為個別考慮 channels 的資訊對其進行卷積之後再將其融合，將 channels 與空間分開是 Extreme Inception 的核心概念。

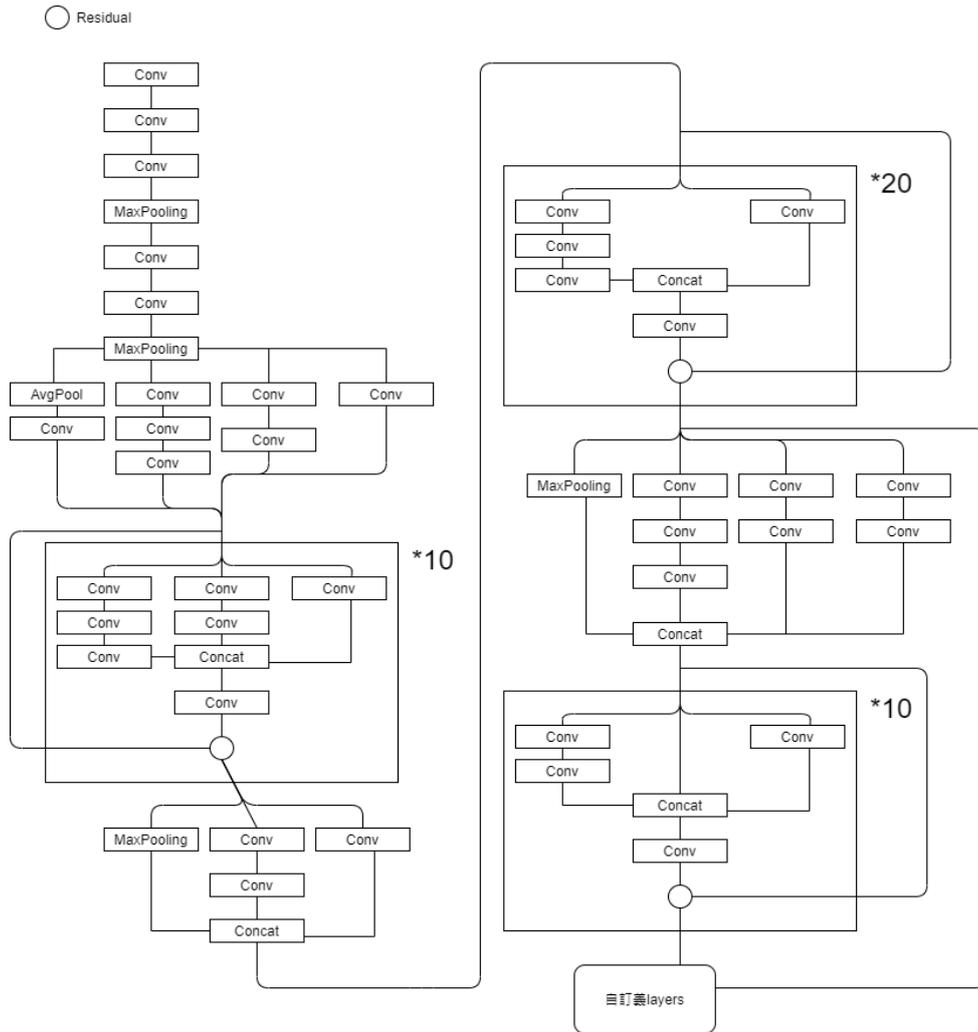
3.2 InceptionV3



圖十五、InceptionV3 架構圖

上圖為 Inception V3 的架構，相較於 VGG16 其參數量較低且準度較高，其中 inception 為本架構的特色，使用較小的 filter 可以有效的降低所需的計算量，如 5*5 的 filter 計算量為 3*3 的 filter 的 25/9 倍，因此 Inception V3 可以實現快速的訓練。同時較寬的網路結構也可以避免產生瓶頸導致重要參數遺失，讓模型可以順利的擷取特徵有關的參數，在本案例中 將模型的 top layers 移除之後接上自訂義的輸出層以符合本次使用數據預測的用途。

3.3 InceptionResNetV2



圖十六、InceptionResNetV2 架構圖

上圖為 InceptionResNet V2 的架構，藉由增加 residual 的運算，使得模型深度可以加深，讓模型參數可以增加，屬於較大型的模型層數達到 572 層之譜，是本次嘗試模型中層數最高的模型，因為使用 inception 的結構參數使用效率較佳，所以模型所佔用記憶體空間還是比深度 23 層的 VGG16 小上許多。

表八、三種模型與 VGG16 比較表

Model	size	top-1 accuracy	top-5 accuracy	Parameters	depth
VGG16	528MB	0.713	0.901	138357544	23

Xception	88MB	0.79	0.945	22910480	126
InceptionV3	92MB	0.779	0.937	23851784	159
InceptionResNetV2	215MB	0.803	0.953	55873736	572

上表為本次使用的模型與 VGG16 的簡單比較，準確度為使用 ImageNet validation dataset 測試之結果，本實驗會將模型的 Top Layers 移除並增加自訂的輸出層以符合預測年齡的目標。

4. 超參數調整與模型績效比較

4.1 擴增前資料 初步模型超參數調整

我們由擴增前的資料集，以下列三種不同模型架構的模型進行訓練與驗證，可以得到的結果如下表：

表九、模型績效比較表

Model ID	X1	In1	inR1
Pre-trained Model	Xception_1	InceptionV3	InceptionResNetV2
模型架構	Xception +GlobalMaxPooling2D +Flatten +Dense(10) +Dense(output)	Inception +GlobalMaxPooling2D +Flatten +Dense(10) +Dense(output)	InceptionResNetV2 +GlobalMaxPooling2D +Flatten +Batchnormalization(momentum=0.8) +Dense(10) +Dense(output)
Activation of dense layer	relu		
Optimizer	adam		
Loss	mse(mean_squared_error)		
epoch	30		
Test MAE	12.41	38.93	54.55

接下來我們針對三種模型，依照不同模型架構特性，進行進一步超參數調整後，找出最佳表現模型。下表列出所有測試模型中最佳的 9 組超參數設計(根據表八中的模型架構，僅列出因實驗設計有差異的參數)，調整的超參數包含:

- (1) BatchNorm: 是否在 Flatten layer 後使用 Batch Normalization
- (2) Optimizer: 使用 Adam 或 Adadelta
- (3) Output layer activation: 輸出層使用的激活函數，使用 Linear 或是 Leaky Relu
- (4) Loss function: 損失函數定義，使用 MSE (mean squared error)或是 MAE (mean absolute error)

Model ID	X1	X2	X3
BatchNorm Optimizer Loss function	None Adam mse	None Adadelta mae	Yes Adadelta mae
Test MAE	12.41	28.05	14.12

Model ID	In1	In2	In3
BatchNorm Output layer activation	None linear	Yes linear	Yes leaky Relu
Test MAE	38.93	16.1	16.8

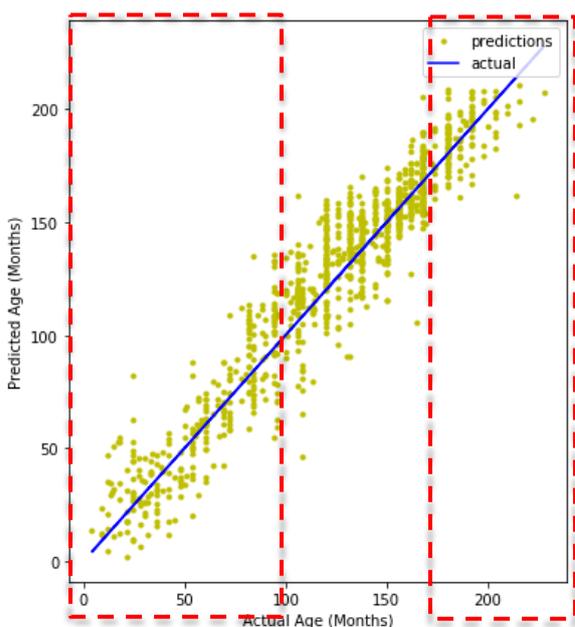
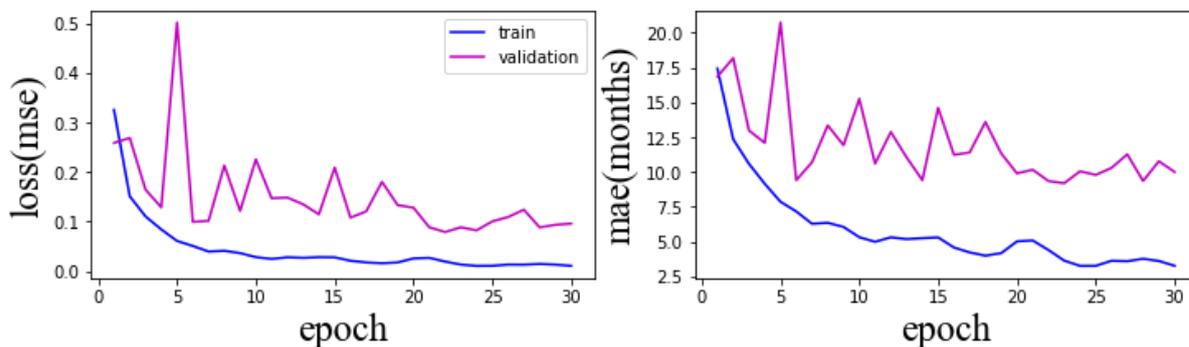
Model ID	InR0	InR1	InR2
BatchNorm Optimizer	None Adam	Yes Adam	Yes Adadelta
Test MAE	286.32	54.55	16.8

圖 十七、擴增前初步模型架構績效

由我們選擇出最佳模型組合為使用 Xception 架構之 X1 模型，後續將做更詳細的分析。

Layer (type)	Output Shape	Param #
xception (Model)	(None, 8, 8, 2048)	20861480
global_max_pooling2d_4 (Glob	(None, 2048)	0
flatten_4 (Flatten)	(None, 2048)	0
dense_10 (Dense)	(None, 10)	20490
dense_11 (Dense)	(None, 1)	11
=====		
Total params: 20,881,981		
Trainable params: 20,827,453		
Non-trainable params: 54,528		

圖 十八、X1 模型架構圖



經實際驗證，使用未擴增資料集(如下圖分佈)訓練後的模型預測結果見左圖。
 可以發現在年齡小於 100 個月與年齡大於 175 個月預測較分散且偏差較大。

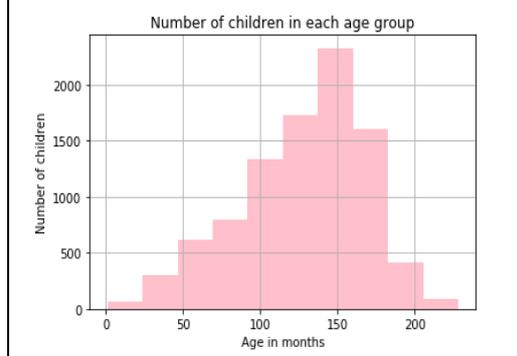
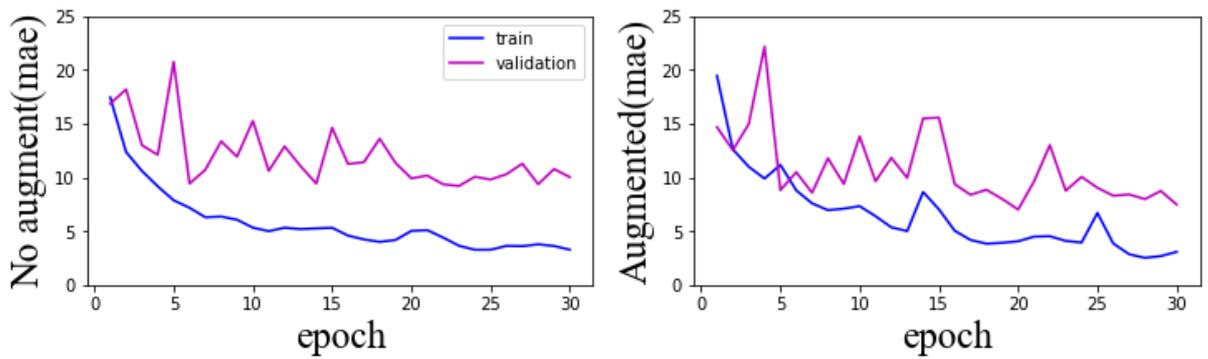


圖 十九、X1 模型訓練過程

4.2 擴增後資料 模型超參數調整

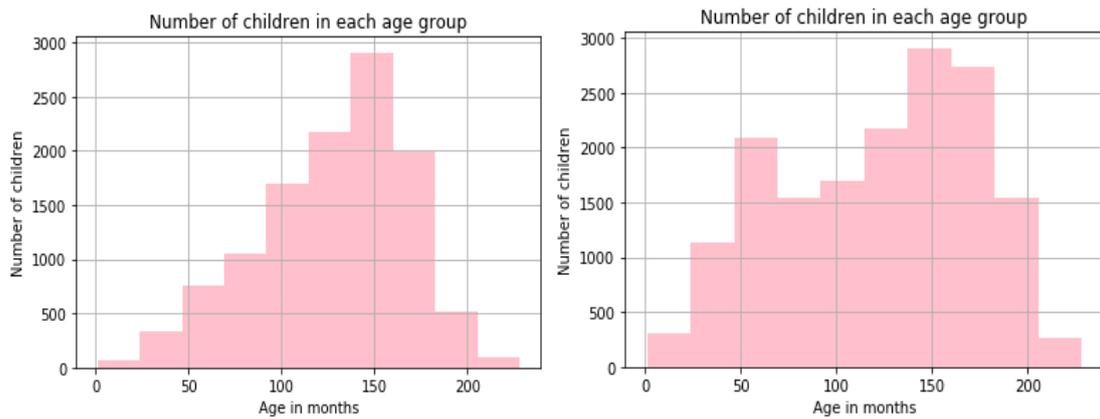
下表使用模型 x_1 比較使用擴增前與擴增後資料集之績效，可以觀察到年齡小於 100 與年齡大於 175 的預測結果，在擴增訓練集後較實際年齡集中，由此可知資料增強後的模型訓練確實可以得到較好的結果，**績效提升 13.53%**。



圖二十、擴增前(左圖)與擴增後(右圖)模型訓練過程比較(均使用 X_1 模型)

表十、資料擴增前後測試集預測結果比較

	擴增前	擴增後
預測結果與實際值比較圖		
測試集 MAE	12.41	10.73



圖二十一、原訓練集年齡分布(左圖)與增強後訓練集年齡分佈(右圖)

在 4.1 節中使用擴增前資料訓練得出模型 x_1 有最佳績效，此階段使用擴增後資料並基於 x_1 模型架構(更名為 x_{a1} 以區分)做進一步調整，比較各模型超參數組合績效。在初步選定模型時，已嘗試出超參數 Optimizer 使用 Adam 與 Loss function 使用 mse 為較佳的超參數組合。

以下模型皆使用固定之超參數：

- (1) Optimizer
- (2) Loss function

所調整的超參數有：

- (1)Dense layer 數量: 連接 1 層或 2 層
- (2)Dense layer 的激活函數: 使用 Relu 或 Sigmoid
- (3) BatchNorm: 是否在 Flatten layer 後使用 Batch Normalization

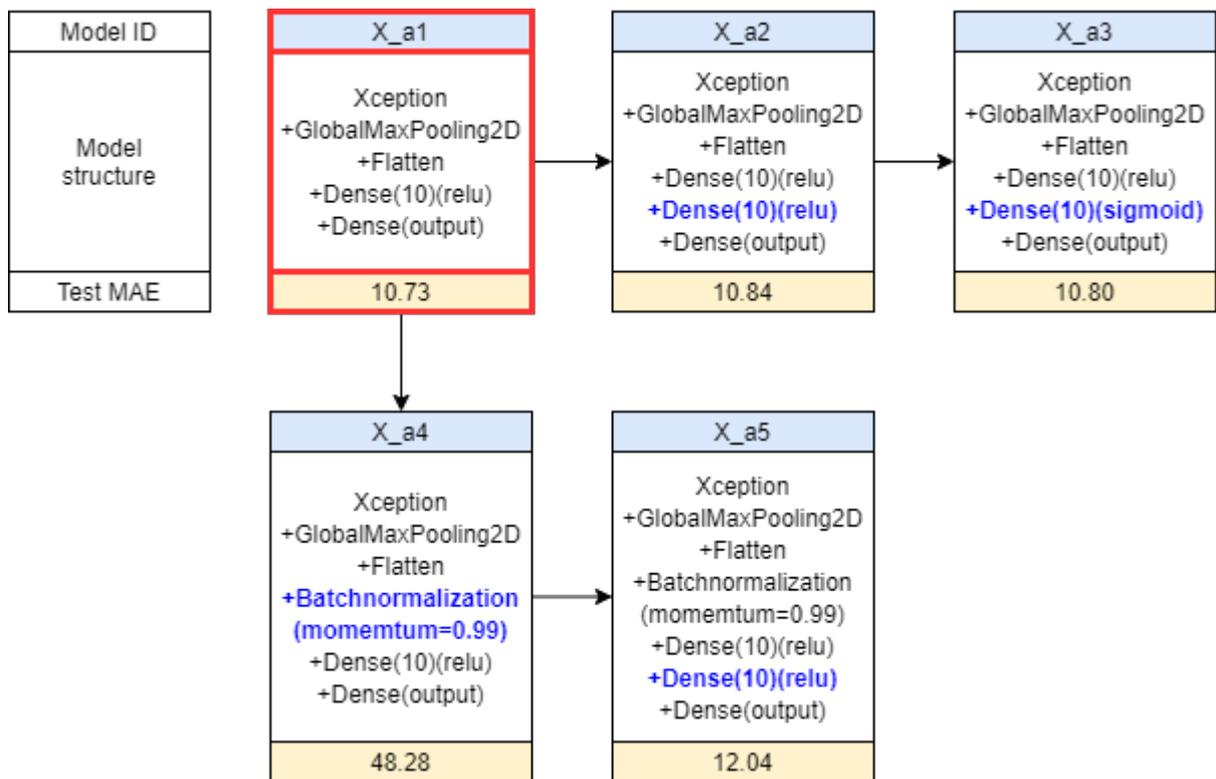


圖 二十二、擴增後資料模型績效

可以觀察到增加一層 Dense 對模型績效影響不大，增加 Batchnormalization 則會使績效變差。兩種方法均無法提升績效，X_a1 模型仍為最佳。

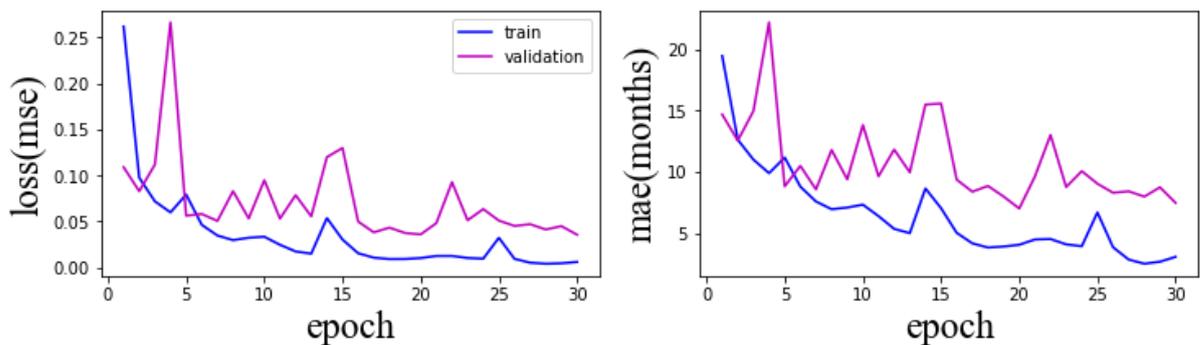


圖 二十三、X_a1 模型訓練過程

三、結果與討論

從模型結果可以得出，在嘗試了上述不同模型後，可以得知本研究中 X_a1 模型在手骨年齡的識別結果最好，精準度可以達到 MAE 10.73 個月，表示目前計畫模型對於骨骼年齡的辨識可以控制誤差在 1 年以內。請教相關專家後，專家解釋手骨能夠用來判斷年齡的依據是在於骨頭內的生長板的發育，主要適用於青年以下的年齡判定。

針對骨骼年齡判定，若有更多的相關資料能夠加入新的資料集，想必更能達成更高的精準度。

四、未來展望

透過 CNN 來輔助手骨年齡判定，可有效協助刑案現場人員進行判斷，而在考古學界，也能夠加速遺址現場的人骨分析。未來，若能增加其他訓練資料類型，例如，其他的人骨部位或是其他物種的骨骼，將能夠更有效的協助進行現場骨骼的判斷。由於現在訓練的資料集是來自 X 光片，若是日後可以加上實體照片的訓練資料，並且結合穿戴裝置(例如眼鏡)，將更有效的協助現場人員直接在現場進行年齡判定，可直接在現場進行判別分類，將大幅提升現場人員的作業效率。

五、參考資料

1. 資料來源：<https://www.kaggle.com/kmader/rsna-bone-age>
2. 深度學習—卷積神經網路的經典網路:
<https://www.itread01.com/content/1549462862.html>
3. 輕量化卷積神經網路: <https://www.itread01.com/content/1550288179.html>
4. Inceptionv3: <https://en.wikipedia.org/wiki/Inceptionv3>
5. 借助殘差網路進一步提升圖像分類水準: <https://kknews.cc/tech/e4vr94.html>
6. Simon Mays, 2010, The Archaeology of Human Bones