Using LLMs to OCR handwriting on manuscript collections and messy WW2 forms

Kent Fitch

UNSW@Canberra & Project Computing Pty Ltd

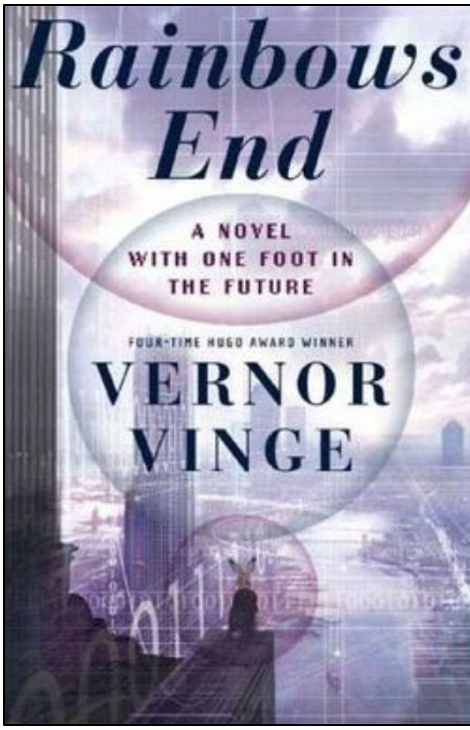
kent.fitch@projectComputing.com

AI4LAM, Canberra, 2 Dec 2025

Three tiny topics

1. Handwriting recognition has been solved by open models — 5 min
2. But be aware when processing large images with small/cheap models — 3 min
3. LoRA fine-tuning a small QWEN VL to improve reading handwritten forms — 7 min

1. Using a small, open-weight LLM to read handwriting



Rainbows End, a 2006 sci-fi novel set mostly in SanDiego - that's the Geisel Library at the university, named after Theodore Seuss Geisel - "Dr Seuss".

Slightly interesting utopian/dystopian prediction of the future with ubiquitous distributed computing and networking, set in 2025 - now! It contained a description of a radical method of book scanning, probably triggered by the new and high profile Google Books project, that stopped me in my tracks, as I was at the time obsessing over optimising digital newspaper scans for the NLA's newspapers project, which morphed into & formed the cornerstone of Trove…

Navicloud custom debinder

Robert could see two white-suited figures, their jackets labeled "Huertas Data Rescue" ... They might have been construction workers. In fact, this business was the ultimate in deconstruction first one and then the other would pull books off the racks and toss them into the shredder's maw … a "NaviCloud custom debinder". The fabric tunnel that stretched out behind it was a "camera tunnel"…

The shredded fragments of books and magazines flew down the tunnel like leaves in tornado, twisting and tumbling. The inside of the fabric was stitched with thousands of tiny cameras. The shreds were being photographed again and again, from every angle and orientation, till finally the torn leaves dropped into a bin just in front of Robert. Rescued data.

This vision was disturbing yet appealing.

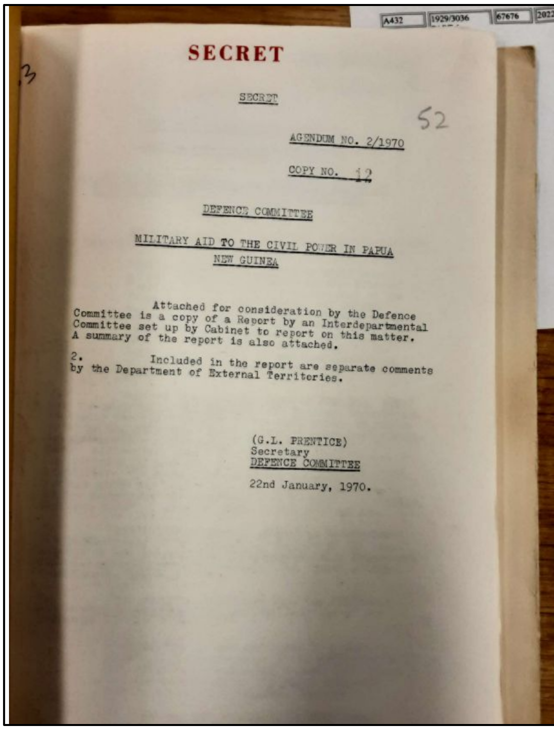

```
[ 20241118_163725.jpg ]

NATIONAL LIBRARY OF AUSTRALIA
Pictures & Manuscripts Branch
MS 882
Papers of Edmund Leolin Piesse
Series 4-5
Box 3
bags, food
ink
```

So here we are, almost 20 years later, living in the future. A few months ago, an academic at the ANU handed me a thumb-drive with over 1300 images he'd taken with his phone when looking at manuscript collections at NLA and SLNSW. He wanted to be able to search the text in the photos.

He wasnt as freestyle as the Navicloud Custom Debinder, but he also didnt bother with any received wisdom on How To Digitise Images. Good for him.

So I formatted these as compressed JPGs - a typical image: 1536 x 2048 pixels about 300KB.
Then I used qwen2.5-vl-32b-instruct, an open weight Chinese model from early this year, running on Alibaba Cloud. It cost $4.70 to process all images. Thats 2.8 images per cent or 1M images for $3600.

The extracted text is shown here - it happily OCRs the box label and even the sign behind it just as eagerly.


SECRET

SECRET

52

AGENDUM NO. 2/1970

COPY NO. 12

DEFENCE COMMITTEE

MILITARY AID TO THE CIVIL POWER IN PAPUA NEW GUINEA

Attached for consideration by the Defence Committee is a copy of a Report by an Interdepartmental Committee set up by Cabinet to report on this matter. A summary of the report is also attached.

2. Included in the report are separate comments by the Department of External Territories.

(G.L. PRENTICE)
Secretary
DEFENCE COMMITTEE

22nd January, 1970.

[20220719_103822.jpg]

SECRET
AGENDUM NO. 2/1970
COPY NO. 12

DEFENCE COMMITTEE

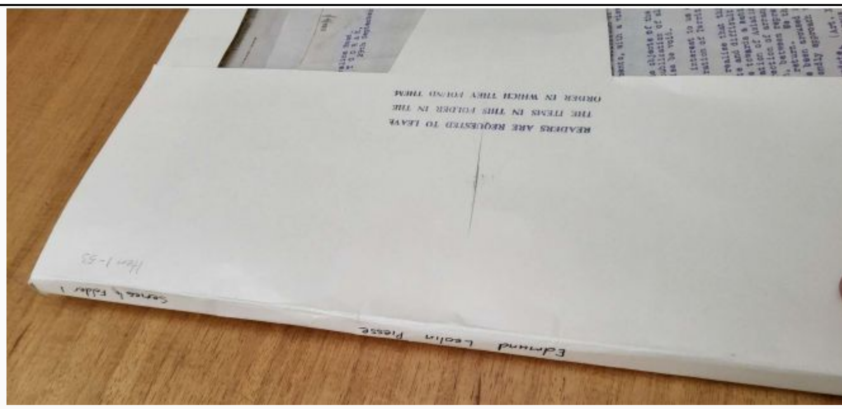
MILITARY AID TO THE CIVIL POWER IN PAPUA NEW GUINEA

Attached for consideration by the Defence Committee is a copy of a Report by an Interdepartmental Committee set up by Cabinet to report on this matter. A summary of the report is also attached.

2. Included in the report are separate comments by the Department of External Territories.

(G.L. PRENTICE)
Secretary
DEFENCE COMMITTEE
22nd January, 1970.

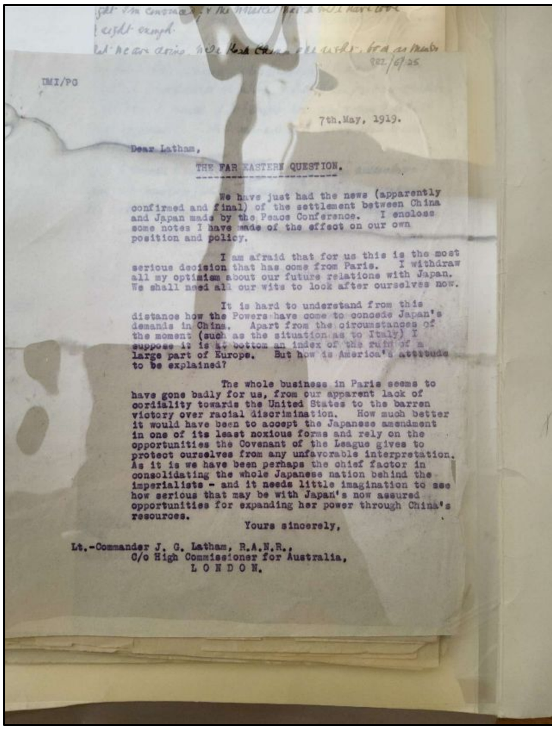
As you'd expect, it does well with type, even if the lighting is random and things are a little skewed.



[20241118_163735.jpg]

READERS ARE REQUESTED TO LEAVE
THE ITEMS IN THIS FOLDER IN THE
ORDER IN WHICH THEY FOUND THEM
Edmund Lealin Please
Series 4 Feller 1
Hem 1-53

Turns out it doesnt care about skew or even rotation at all - it just sucks everything up.



[20241118_164850.jpg]

7th. May, 1919.
Dear Latham,
THE FAR EASTERN QUESTION.

We have just had the news (apparently confirmed and final) of the settlement between China and Japan made by the Peace Conference. I enclose some notes I have made of the effect on our own position and policy.

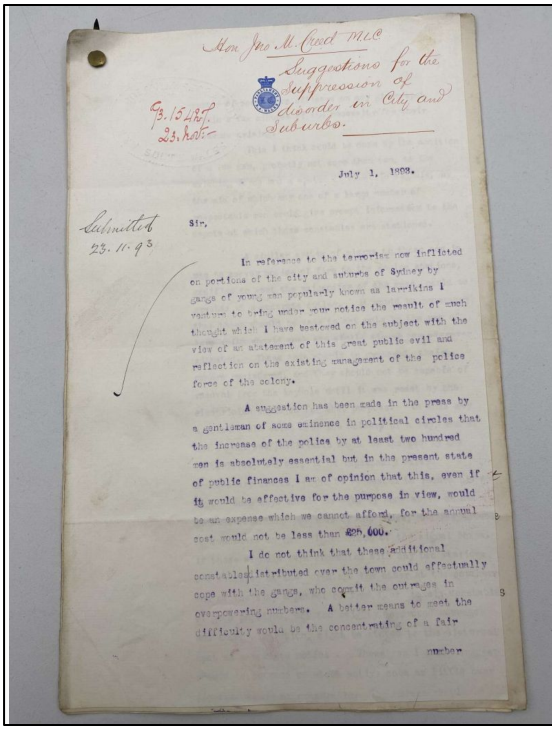
I am afraid that for us this is the most serious decision that has come from Paris. I withdraw all my optimism about our future relations with Japan. We shall need all our wits to look after ourselves now.

It is hard to understand from this distance how the Powers have come to concede Japan's demands in China. Apart from the circumstances of the moment (such as the situation as to Italy), I suppose it is at bottom an index of the ruin of a large part of Europe. But how is America's attitude to be explained?

The whole business in Paris seems to have gone badly for us, from our apparent lack of cordiality towards the United States to the barren victory over racial discrimination. How much better it would have been to accept the Japanese amendment in one of its least noxious forms and rely on the opportunities the Covenant of the League gives to protect ourselves from any unfavorable interpretation. As it is we have been perhaps the chief factor in consolidating the whole Japanese nation behind the imperialists - and it needs little imagination to see how serious that may be with Japan's now assured opportunities for expanding her power through China's resources.

Yours sincerely,
Lt.-Commander J. G. Latham, R.A.N.R.,
C/o High Commissioner for Australia,
L O N D O N.

And certainly doesnt angst about image quality



Hon Jno M. Creed MLC
Suggestions for the Suppression of disorder in City and Suburbs.

93.15427
23. Nov

July 1, 1898.

Submitted
23. 11. 98

Sir,

In reference to the terrorism now inflicted on portions of the city and suburbs of Sydney by gangs of young men popularly known as larrikins I venture to bring under your notice the result of much thought which I have bestowed on the subject with the view of an abatement of this great public evil and reflection on the existing management of the police force of the colony.

A suggestion has been made in the press by a gentleman of some eminence in political circles that the increase of the police by at least two hundred men is absolutely essential but in the present state of public finances I am of opinion that this, even if it would be effective for the purpose in view, would be an expense which we cannot afford, for the annual cost would not be less than £25,000.

I do not think that these additional constables distributed over the town could effectually cope with the gangs, who commit the outrages in overpowering numbers. A better means to meet the difficulty would be the concentrating of a fair

number

[IMG_2186.JPG]

Alan Jao M. Creed M.C.
Suggestions for the
Suppression of
Disorder in City and
Suburbs.

July 1, 1898.

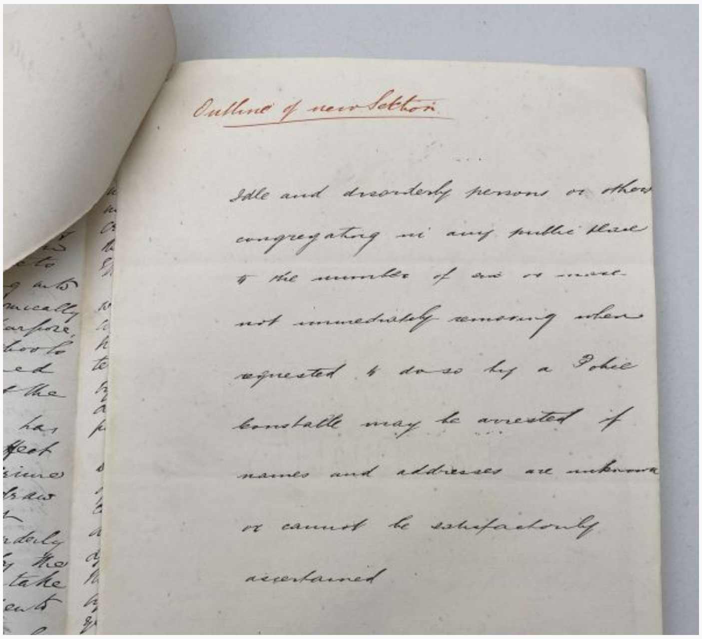
Sir,

In reference to the terrorism now inflicted on portions of the city and suburbs of Sydney by gangs of young men popularly known as larrikins I venture to bring under your notice the result of much thought which I have bestowed on the subject with the view of an abatement of this great public evil and reflection on the existing management of the police force of the colony.

A suggestion has been made in the press by a gentleman of some eminence in political circles that the increase of the police by at least two hundred men is absolutely essential but in the present state of public finances I am of opinion that this, even if it would be effective for the purpose in view, would be an expense which we cannot afford, for the annual cost would not be less than £25,000.

I do not think that these additional constables distributed over the town could effectively cope with the gangs, who commit the outrages in overpowering numbers. A better means to meet the difficulty would be the concentrating of a fair number

And has no problem with pages containing both handwriting and text, even faint text.


Outline of new Section

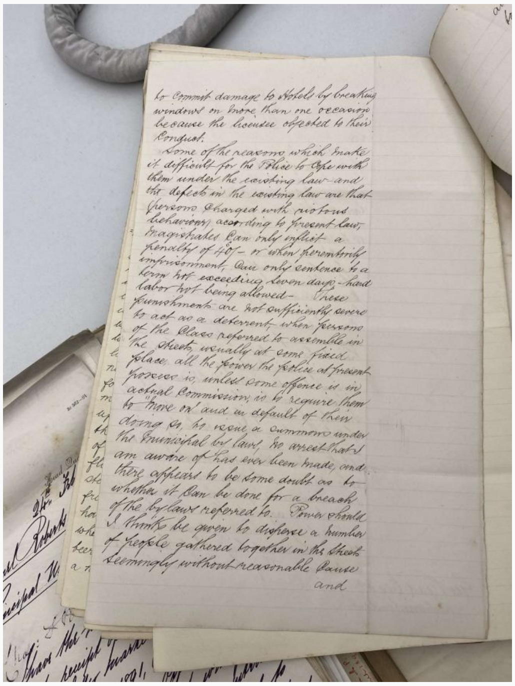
Idle and disorderly persons or others
congregating in any public place
to the number of two or more
not immediately removing when
requested to do so by a Police
constable may be arrested if
names and addresses are unknown
or cannot be satisfactorily
ascertained

[IMG_2182.JPG]

Outline of new Section

Idle and disorderly persons or others congregating in any public place to the number of two or more not immediately removing when requested to do so by a Police constable may be arrested if names and addresses are unknown or cannot be satisfactorily ascertained.

Handwriting - also no problems!



to Commit damage to Hotels by breaking windows on more than one occasion because the licensee objected to their Conduct.

Some of the reasons which make it difficult for the Police to Cope with them under the existing law and the defects in the existing law are that persons Charged with riotous behaviour, according to present law, Magistrates Can only inflict a penalty of 40/- or when peremtorily imprisonment, Can only sentence to a term not exceeding seven days - hard labor not being allowed - These punishments are not sufficiently severe to act as a deterrent, when persons of the Class referred to assemble in the street, usually at some fixed place, all the power the police at present possess is, unless some offence is in actual Commission, is to require them to "move on" and in default of their doing so, to issue a summons under the Municipal by laws, no arrest that I am aware of has ever been made, and there appears to be some doubt as to whether it Can be done for a breach of the bylaws referred to. Power should I think be given to disperse a number of people gathered together in the street seemingly without reasonable Cause

and

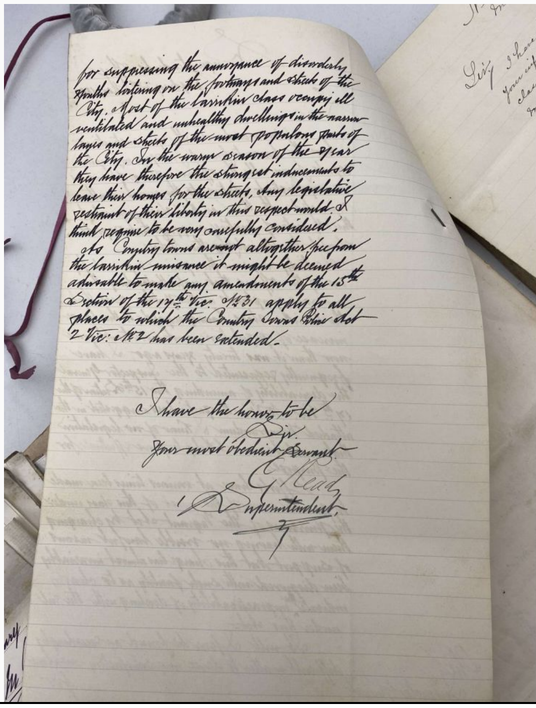
[IMG_2172.JPG]

to commit damage to hotels by breaking windows on more than one occasion because the licensee objected to their conduct.

Some of the reasons which make it difficult for the police to deal with them under the existing law— and the defects in the existing law are that persons charged with riotous behaviour, according to present law, magistrates can only inflict a penalty of 40s. or when permanently imprisoned, can only sentence to a term not exceeding seven days' hard labour—but being allowed—these punishments are but sufficiently severe to act as a deterrent when persons of the class referred to assemble in the street, usually at some fixed place, all the powers the police officer has is unless some offence is in actual commission, is to require them to move on and in default of their doing so, he issues a summons under the Municipal law. In an event that I am aware of has ever been made, and this affairs to be done doubt as to whether it can be done for a breach of the bylaws referred to. Power should I think be given to disperse a number of people gathered together in the street seemingly without reasonable cause and

Even when dense, mostly good, but, for example:

all the powers the police at present possses
OCRed as **all the powers the police officer has**


for suppressing the annoyance of disorderly youths loitering on the footways and streets of the City. Most of the larrikin class occupy ill ventilated and unhealthy dwellings in the narrow lanes and streets of the most populous parts of the City. In the warm season of the year they have therefore the strongest inducements to leave their homes for the streets. Any legislative restraint of their liberty in this respect would, I think, require to be very carefully considered.

As Country towns are not altogether free from the larrikin nuisance it might be deemed advisable to make any amendments of the 15th Section of the 17th Vic. No. 31 apply to all places to which the Country Towns Police Act 2 Vic: No. 2 has been extended.

I have the honor to be
Sir
Your most obedient Servant
Superintendent

[IMG_2149.JPG]

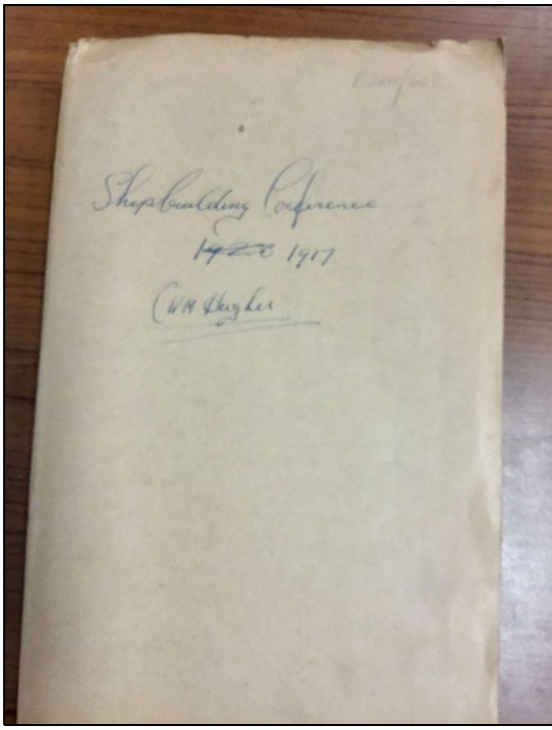
for suppressing the annoyance of disorderly people lurking on the thoroughfares and streets of the City, a part of the lower classes very ill regulated and unhealthy dwelling in the narrow lanes and courts of the most populous parts of the City. In the warm season of the year they have therefore the strongest inducement to leave their homes for the streets. Any legislative restraint of their liberty in this respect would, I think, require to be very carefully considered.

As Country towns are quite altogether free from the nuisance it might be deemed advisable to make any amendments of the 15th section of the 14th Act apply to all places to which the Country Sessions Act 2 Vic. c. 72 has been extended.

I have the honor to be
Your most obedient servant,
[Signature]

And very stylised handwriting is fine too, but again the odd error:

Loitering OCRed as **lurking**
Larrikin classes OCRed as **lower classes**



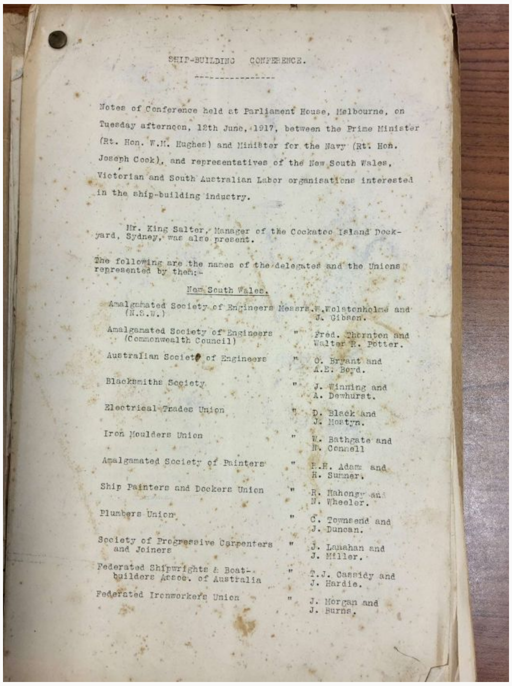
[20241112_110712.jpg]

Shipbuilding Conference

1928 1917

W.M. Hughes

It tolerates out of focus content



[20241112_110722.jpg]

SHIP-BUILDING CONFERENCE.

Notes of Conference held at Parliament House, Melbourne, on Tuesday afternoon, 12th June, 1917, between the Prime Minister (Rt. Hon. W.M. Hughes) and Minister for the Navy (Rt. Hon. Joseph Cook), and representatives of the New South Wales, Victorian and South Australian Labor organisations interested in the ship-building industry.

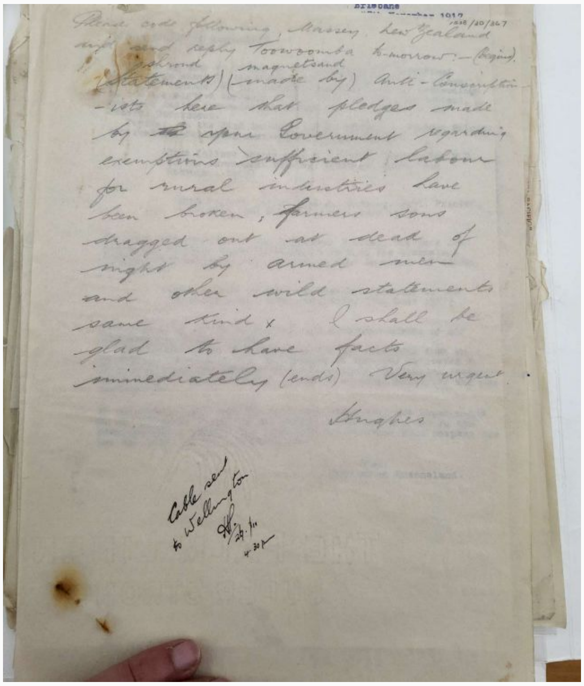
Mr. King Salter, Manager of the Cockatoo Island Dockyard, Sydney, was also present.

The following are the names of the delegates and the Unions represented by them:-

New South Wales.

Amalgamated Society of Engineers (N.S.W.) - Messrs. W. Wolstonholme and J. Gibson.
Amalgamated Society of Engineers (Commonwealth Council) - Fred. Thornton and Walter R. Potter.
Australian Society of Engineers - Q. Bryant and A.E. Boyd.
Blacksmiths Society - J. Winning and A. Dewhurst.
Electrical Trades Union - D. Black and J. Mostyn.
Iron Moulders Union - W. Bathgate and N. Connell.
Amalgamated Society of Painters - P.H. Adam and H. Summer.
Ship Painters and Dockers Union - R. Mahony and N. Wheeler.
Plumbers Union - C. Townsend and J. Duncan.
Society of Progressive Carpenters and Joiners - J. Luahan and J. Miller.
Federated Shipwrights & Boatbuilders Assoc. of Australia - T.J. Cassidy and J. Hardie.
Federated Ironworkers Union - J. Morgan and J. Burns.

And paper quality imperfections, and the page layout is usefully honoured



[20250627_142744.jpg]

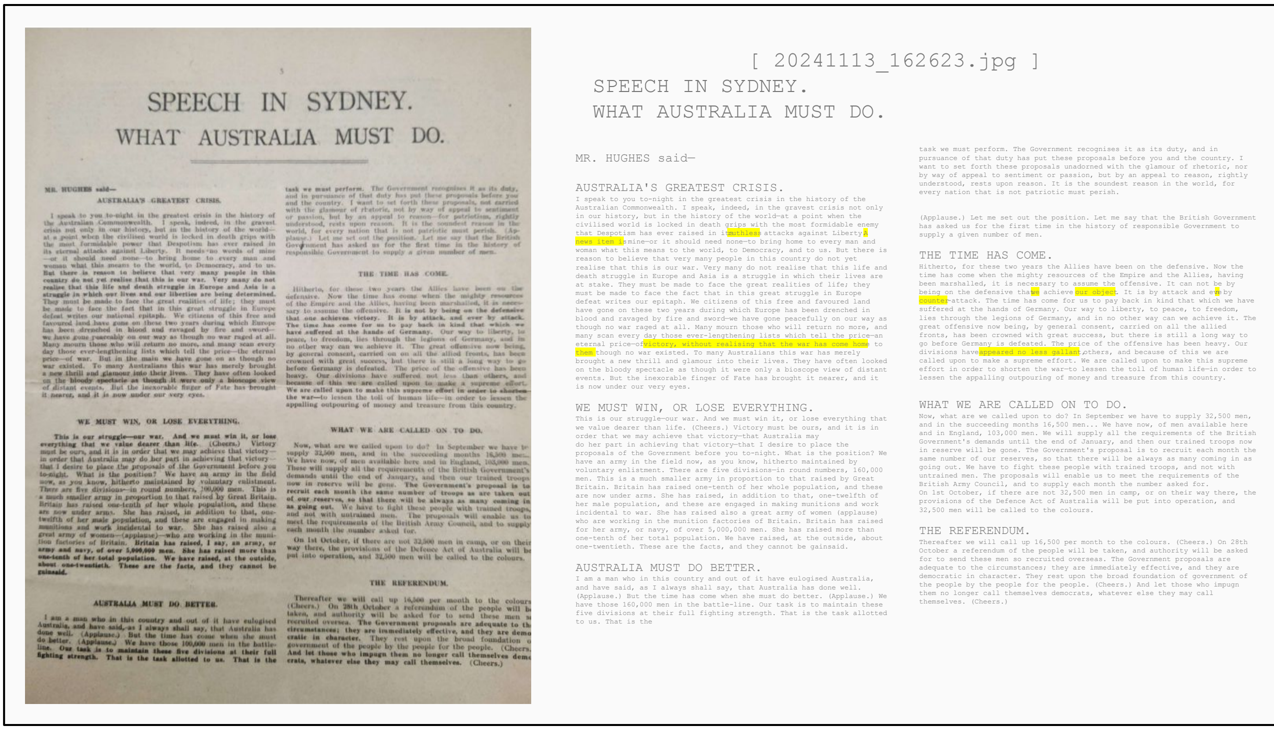
Please send following, Massey, New Zealand and Kelly, Toowoomba to-morrow. (Legends) (made by) Anti-Conscriptionists here that pledges made by the Australian Government regarding exemptions sufficient labour for rural industries have been broken; farmers sons dragged out at dead of night by armed men and other wild statements same kind x I shall be glad to have facts immediately (ends). Very urgent.

Hughes
Cable sent
to Wellington
24th 11
4.30 a.m.

It isnt perfect - here, WM Hughes' scribbled **"and send reply"** becomes **"and Kelly"**.

"**Statement**" becomes **Legends**, and I think, but am not sure, the date should be the 27th not the 24th (based on context pages)

The best thing of course is that for under $5, over 1300 images are now searchable for text and soon, for meaning.



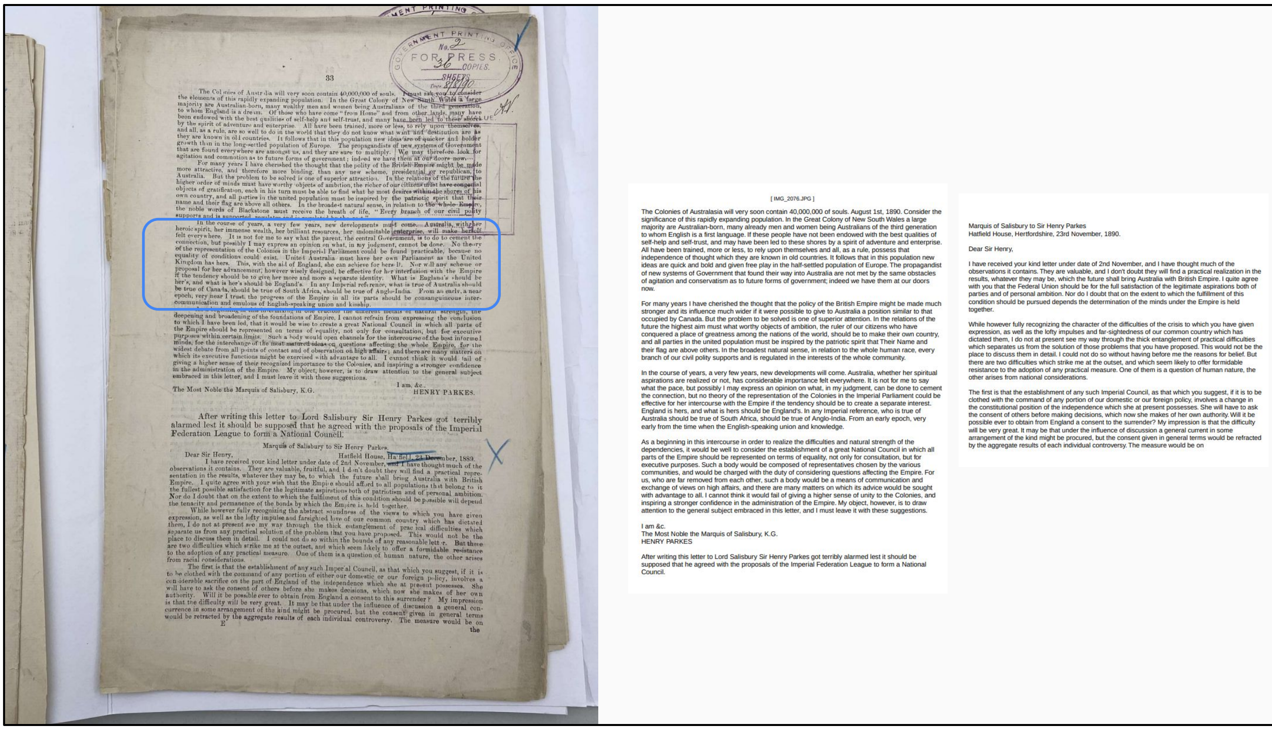
Multi columns are handled well, returning text in the expected reading order.
But on this dense page, there are some unexpected OCR errors:

External OCRed as **ruthless**   a word which looks quite different

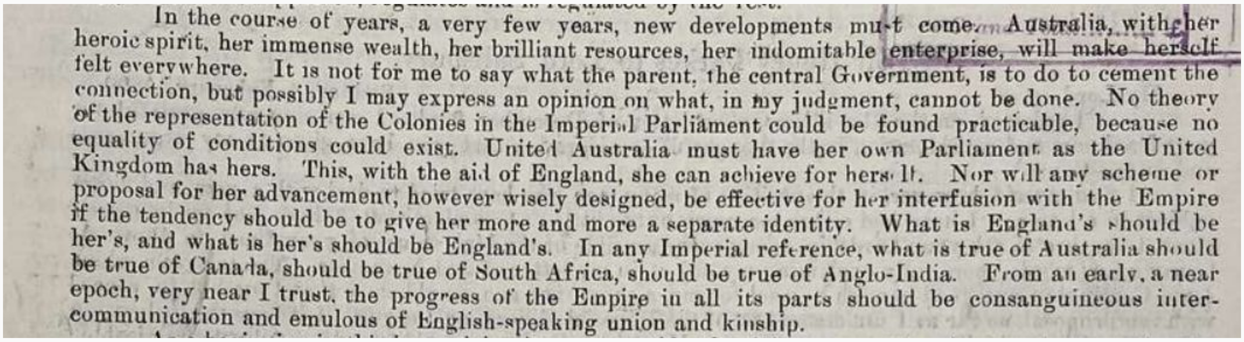
That one achieves victory -> **that we achieve our object**

These errors are NOT typical OCR error: they are less "perception" and more "semantic" induced errors, which brings us to…

2. OCR accuracy degrades in strange ways on large images processed by visual models



Here's a noisy image with a lot of small text.
The top part by Henry Parkes. OCRing generates about 1.5 pages of A4.
The OCR looks plausible - reads well, most of it matches, but lets look at the highlighted para…



In the course of years, a very few years, new developments must come. Australia, with her heroic spirit, her immense wealth, her brilliant resources, her indomitable enterprise, will make herself felt everywhere. It is not for me to say what the parent, the central Government, is to do to cement the connection, but possibly I may express an opinion on what, in my judgment, cannot be done. No theory of the representation of the Colonies in the Imperial Parliament could be found practicable, because no equality of conditions could exist. United Australia must have her own Parliament as the United Kingdom has hers. This, with the aid of England, she can achieve for herself. Nor will any scheme or proposal for her advancement, however wisely designed, be effective for her interfusion with the Empire if the tendency should be to give her more and more a separate identity. What is England's should be her's, and what is her's should be England's. In any Imperial reference, what is true of Australia should be true of Canada, should be true of South Africa, should be true of Anglo-India. From an early, a near epoch, very near I trust, the progress of the Empire in all its parts should be consanguineous inter-communication and emulous of English-speaking union and kinship.

Just looking at the end of top line, we read as "**new developments must come. Australia, with her…**"

Truth

In the course of years, a very few years, new developments **must** come.

Australia**, with her heroic spirit, her immense wealth, her brilliant resources, her indomitable enterprise, will make herself** felt everywhere.

It is not for me to say what the parent, the **Central Government, is to do to cement the connection**, but possibly I may express an opinion on what, in my judgement, **cannot** be done. No theory of the representation of the Colonies in the Imperial Parliament **could be found practicable, because no equality of conditions could exist. United Australia must have her own Parliament as the United Kingdom has hers. This, with the aid of England, she can achieve for herself.** Nor will any scheme or proposal for her advancement, **however wisely designed,** be effective for her **interfusion** with the Empire if the tendency should be to **give her more and more** a separate **identity**. **What is** England**'s should be** her's, and what is her's should be England's. In any Imperial reference, **what** is true of Australia **should be true of Canada,** should be true of South Africa, should be true of Anglo-India. From an early, **a near** epoch, very **near I trust, the progress of the Empire in all its parts should be consanguineous inter-communication and emulous of** English-speaking union and **kinship.**

Truth adjacent

In the course of years, a very few years, new developments **will** come.

Australia, **whether her spiritual aspirations are realized or not, has considerable importance** felt everywhere.

It is not for me to say what the **pace**, but possibly I may express an opinion on what, in my judgment, **can** be done to cement the connection, but no theory of the representation of the Colonies in the Imperial Parliament could be effective for her **intercourse** with the Empire if the tendency should be to **create** a separate **interest**. England is hers, and what is hers should be England's.
In any Imperial reference, **who** is true of Australia should be true of South Africa, should be true of Anglo-India. From an early epoch, very **early from the time when the** English-speaking union and **knowledge.**

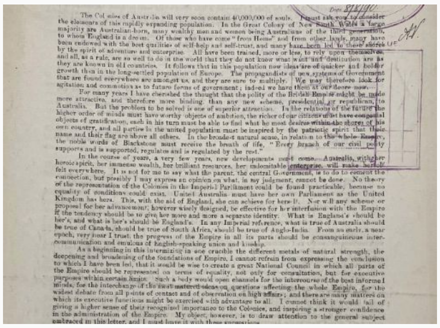
We see the OCR is very poor. What is startling is that the errors arent caused by mistaking letters such as c for an e - a normal OCR engine would never confuse the word "**must**" for "**will**" as they dont even look vaguely similar.

So was it riffing on Henry Parkes embedded model knowledge? No - cutting his name from the image and reOCRing gave a different result, but no better. The problem is much more general than that.

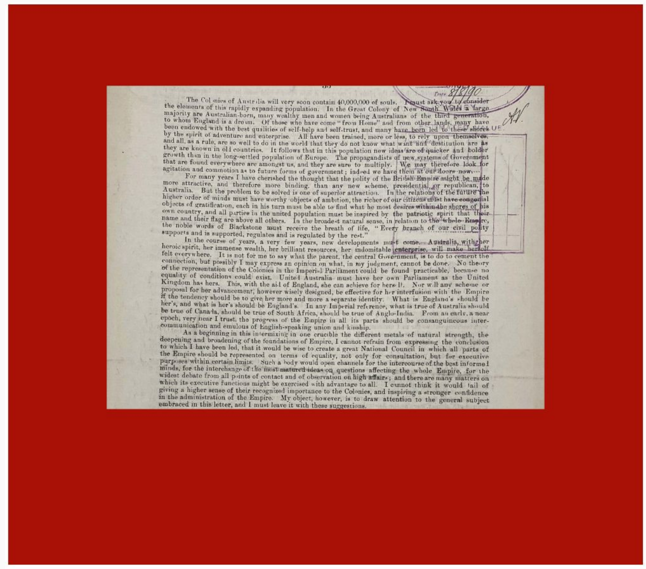
The font is very small at the downsized image - maybe that's the problem.

But if we try the original camera image - 12Megapixels (3024 x 4032), the OCR is even worse!

What is happening reveals something fundamentally surprising about how visual models work.



Crop this section as a 1540 x 2072 pixels image

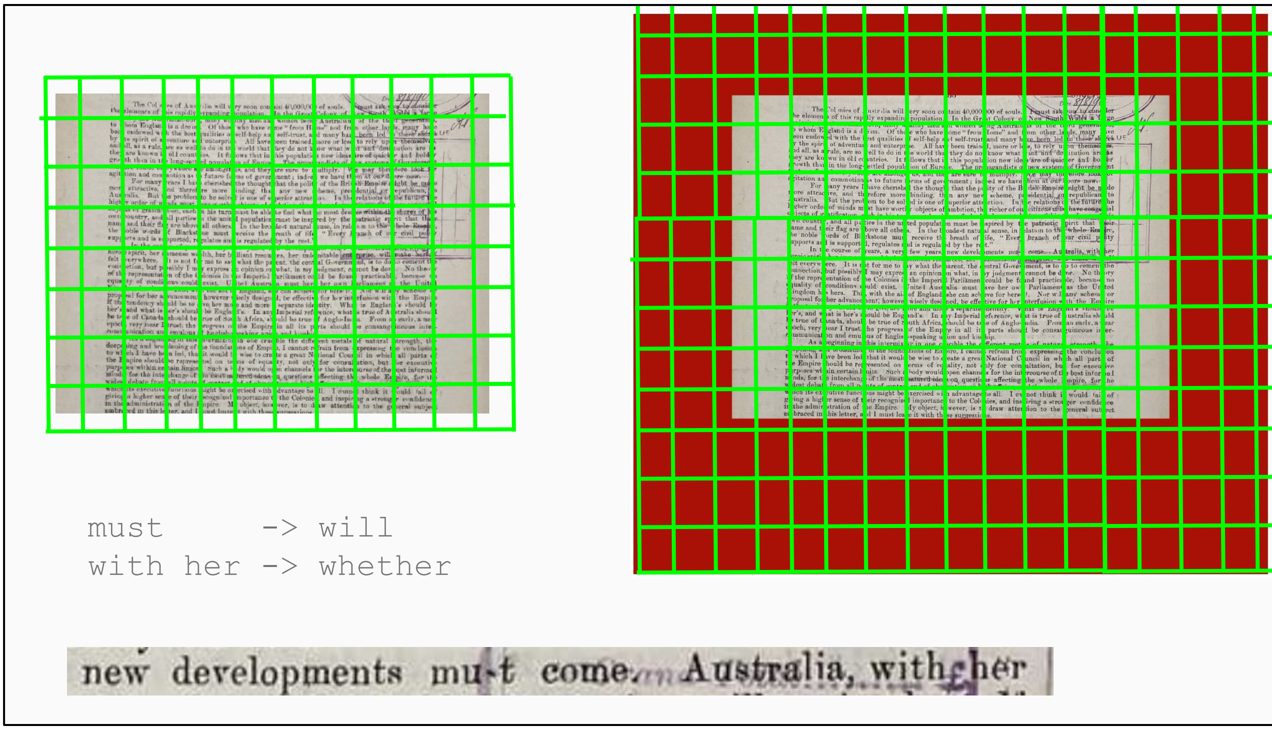


Then embed this section (same res) into 3000 x 3000 image with content-free background

If we crop part of the original camera image to make an image 1540 x 2072 px, as on the left, the OCR is accurate.

But if we take that crop and embed it in a content-free background, and make it 3000 x 3000 px, surely it should still be able to OCR it just the same, right?

But that's not what happens - the OCR on this right hand image is very poor.



To see what's going on, we need to understand how the visual transformer processes images.
In the case of QWEN2.5 VL, it first divides the image patches of 14 x 14 px. These patches can be thought of as the equivalent of text tokens for a normal text stream into a LLM.

However, unlike text tokens, the visual transformer cant simply map a token to an embedding, which represents the token's information, it's 2D position and it's context in a high dimensional space . Instead, converting a series of tokens into a series of embeddings is done dynamically using computationally expensive attention layers in the Visual Transformer.

But a large image will generate more tokens hence embeddings than the model wants to handle. So at some point, depending on configuration, the model will be forced to reduce the pixel resolution of the image - basically to combine or average patches, hence losing information.

In this particular case, the smaller image didnt trigger this downsampling, but a 3000 x 3000 image did, causing the resolution loss that led to the poor OCR.
So how does the word **must** become **will**?

Without downsampling, the embeddings give a very strong hint to the LLM that at some particular location, the pixels making up a visual representation of the word "mu-t" were present. The LLM reliably hallucinates this "mu-t" into "must" given the context. But when downsampling occurs, this hint will be much weaker and vaguer - maybe it will just be "probably 4 letters surrounded by white space, first few letters no ascenders or descenders, first letter is quite wide, third letter - WTF - maybe a dash? last letter has an ascender, probably…"

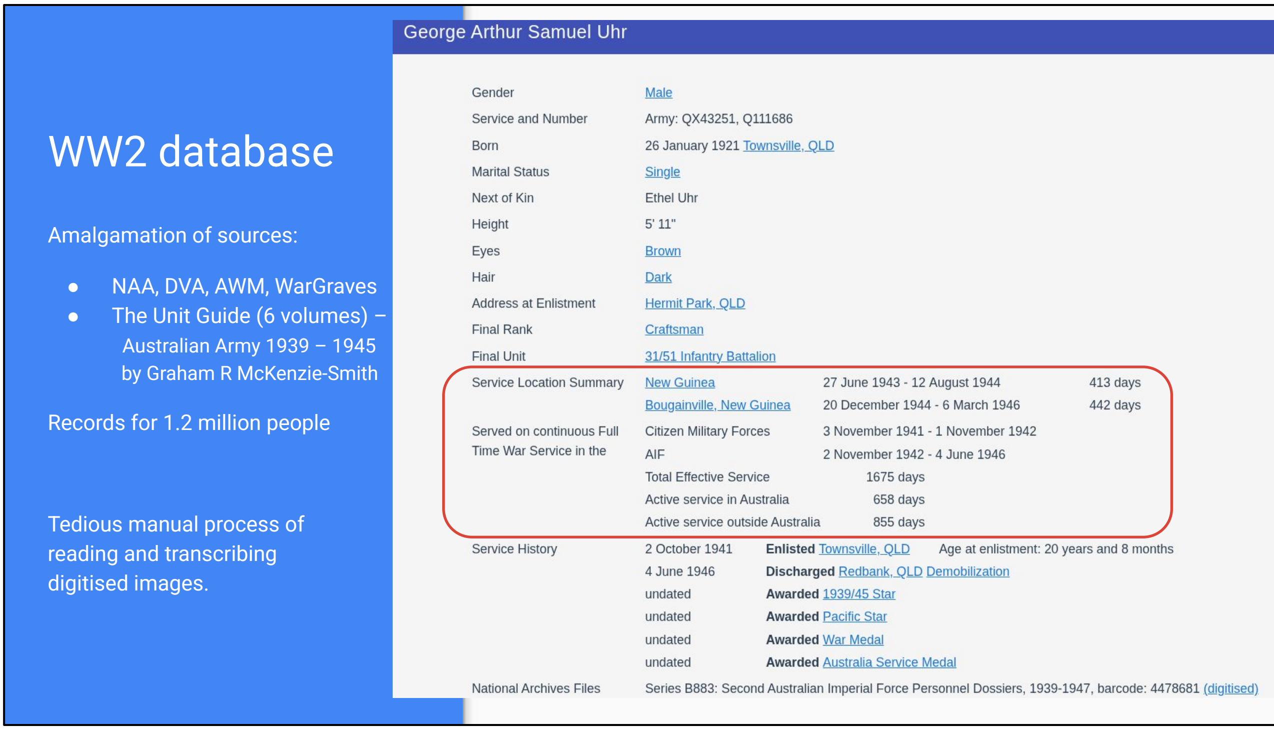
Imagine looking at the text thru dirty glasses, or at a distance - you get the odd word with idiosyncratic shapes and you guess the rest based on that context.

And guessing is exactly what the LLM bolted onto the output of the Visual Transformer excels at - probabilistically picking the next output token based on the previous context and the latest information. So it hallucinates away, giving it's best shot - a failure mode very different from traditional OCR such as Abby Finereader or Tesseract.

The takeaway is that when you find the font size and pixel threshold of the model you're using then it will work well. But beyond that threshold, it will not.

A newer visual model, Deepseek OCR, directly addresses this problem with very effective semantically aware visual compression, but that's another topic..

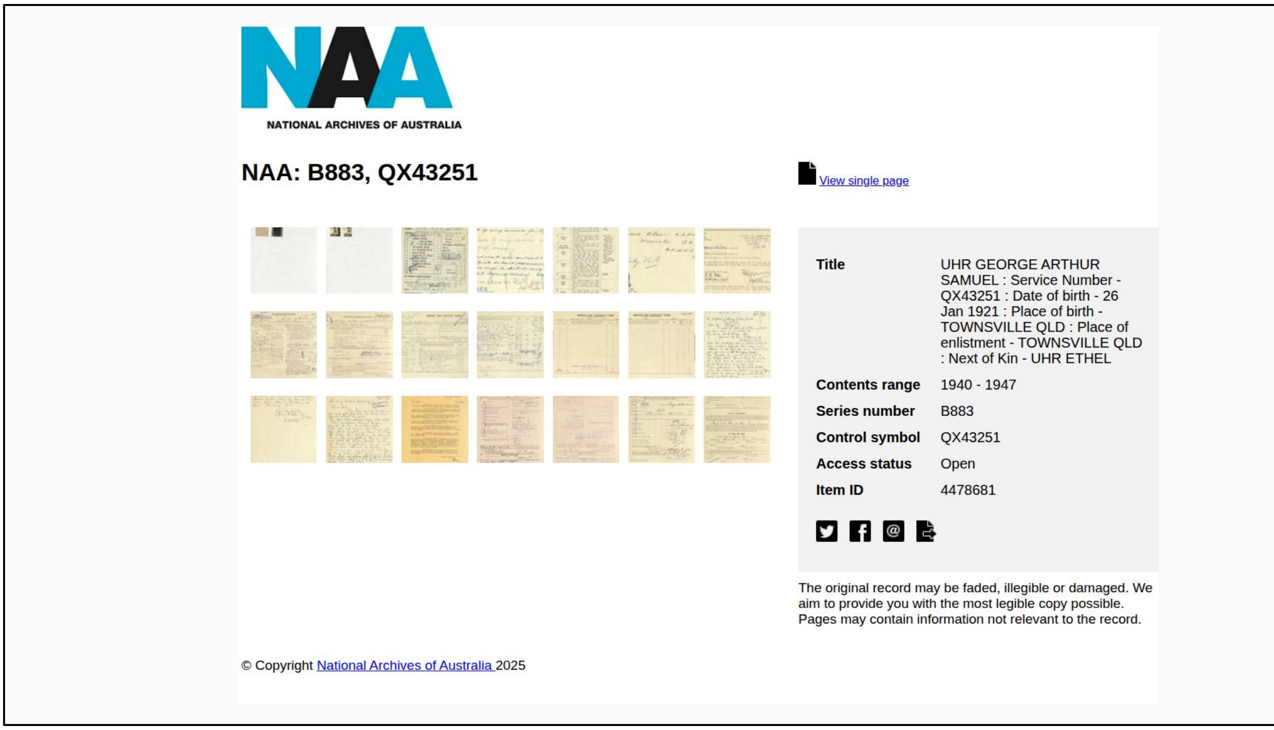
3. LoRA fine-tuning a small QWEN VL to improve reading of handwritten forms



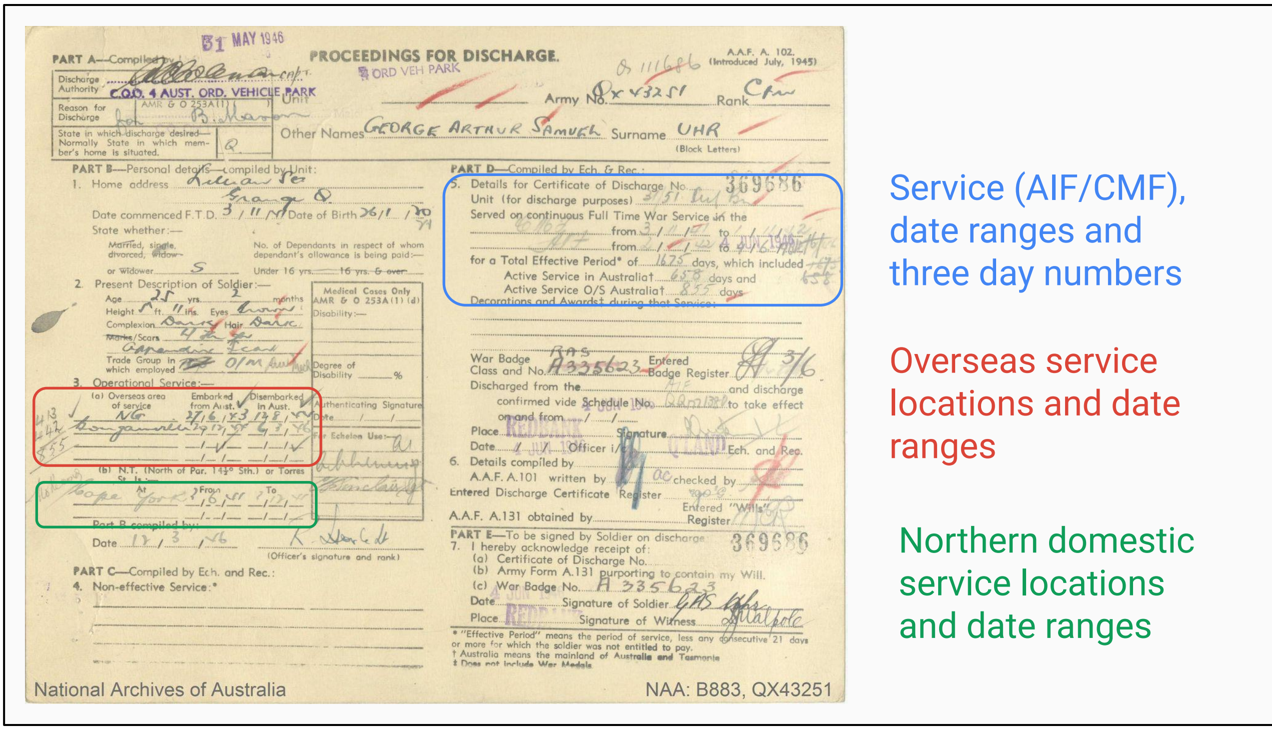
A database documenting the people and units of Australia's military during WW2 is being assembled at UNSW Canberra, following on from their AIF and NZEF databases (WW1), a Vietnam War database and the very large Australians at War video interview project. It currently has around 1.2 million skeletal records that are being fleshed out by a tedious manual processing of reading and transcribing digitised images created by the NAA.

The goal of the WW2 project is to create a resource for researchers and the public that presents as comprehensively as possible, information about Australians who served in WW2 and how they were organised. And we want to do that as soon as possible.

Wondering if we could speed up this process, we're first looking at automating the transcribing of service location and service time summaries, the data highlighted here, captured mostly from service discharge forms.



A discharge form is typically just one of between the 5 and 100 pages in the service file of a typical ww2 enlistee. Almost all have now been digitised by the NAA and are linked from the WW2 database.



31 MAY 1946

PROCEEDINGS FOR DISCHARGE.

A.A.F. A. 102. (Introduced July, 1945)

PART A—Compiled by

Discharge Authority: C.O.O. 4 AUST. ORD. VEHICLE PARK

ORD VEH PARK

Unit

Reason for Discharge: AMR & O 253A(1)( )

State in which discharge desired—Normally State in which member's home is situated. Q

Army No. QX43251 Rank Cfn

Other Names GEORGE ARTHUR SAMUEL Surname UHR (Block Letters)

PART B—Personal details—compiled by Unit:

1. Home address

Date commenced F.T.D. 3/11/41 Date of Birth 26/1/20

State whether:— Married, single, divorced, widow or widower S

No. of Dependants in respect of whom dependant's allowance is being paid:— Under 16 yrs. 16 yrs. & over

2. Present Description of Soldier:— Age 25 yrs. 2 months Height 5 ft. 11 ins. Eyes Brown Complexion Dark Hair Dark Marks/Scars Appendix Scar

Trade Group in which employed

Medical Cases Only AMR & O 253A(1)(d) Disability:— Degree of Disability %

Authenticating Signature Date / /

For Echelon Use:—

3. Operational Service:—

(a) Overseas area of service	Embarked from Aust.	Disembarked in Aust.
NG	27/6/43	12/8/44
	29/12/44	6/3/46
	/ /	/ /
	/ /	/ /

(b) N.T. (North of Par. 14½° Sth.) or Torres St. Is.:—

At	From	To
Cape York	?/6/44	?/12/44
	/ /	/ /

Part B compiled by:

Date 17/3/46

(Officer's signature and rank)

PART C—Compiled by Ech. and Rec.:

4. Non-effective Service:*

PART D—Compiled by Ech. & Rec.:

5. Details for Certificate of Discharge No. 369686

Unit (for discharge purposes) 31/51 Inf Bn

Served on continuous Full Time War Service in the

CMF from 3/11/41 to 1/12/42

AIF from 2/12/42 to 4/6/1946

for a Total Effective Period* of 1675 days, which included

Active Service in Australia† 658 days and

Active Service O/S Australia† 855 days

Decorations and Awards‡ during that Service:

War Badge Class and No. RAS A335623 Entered Badge Register 3/6

Discharged from the AIF and discharge confirmed vide Schedule No. to take effect on and from / /

Place REDBANK Signature

Date 4 JUN 1946 Officer i/c Q'LAND Ech. and Rec.

6. Details compiled by

A.A.F. A.101 written by checked by

Entered Discharge Certificate Register

A.A.F. A.131 obtained by Entered "Wills" Register

PART E—To be signed by Soldier on discharge 369686

7. I hereby acknowledge receipt of:

(a) Certificate of Discharge No.

(b) Army Form A.131 purporting to contain my Will.

(c) War Badge No. A 335623

Date Signature of Soldier

Place REDBANK Signature of Witness

* "Effective Period" means the period of service, less any consecutive 21 days or more for which the soldier was not entitled to pay.

† Australia means the mainland of Australia and Tasmania

‡ Does not include War Medals

National Archives of Australia

NAA: B883, QX43251

Discharge forms come in a few versions, but this is the most common. For now, we're looking to extract information from 3 parts of this form:

1. Service date ranges and days served, broken into 3 categories,
2. Overseas service locations and dates
3 Northern domestic service locations and dates

PART A—Compiled by Unit:

PROCEEDINGS FOR DISCHARGE.

A.A.F. A. 102. (Introduced July, 1945)

2131 days

Discharge Authority	Lanoforces 10184 of 2 Dec 45
Reason for Discharge	AMR & O 253A (1) ( ) AMR 184 A(1) (N) Longer
State in which discharge desired—Normally State in which member's home is situated.	Q'land

Unit 2/187 Aust Sup Depot Pl Army No. QX2641 Rank Pte

Other Names Vernon George Arthur Surname KEILY (Block Letters)

PART B—Personal details—compiled by Unit:

1. Home address 111 Wandal Road ROCKHAMPTON, Q
   Date commenced F.T.D. 6/11/39 Date of Birth 24/6/14
   State whether:— Married, single, divorced, widow or widower Single
   No. of Dependants in respect of whom dependant's allowance is being paid:— Under 16 yrs. 16 yrs. & over 1
2. Present Description of Soldier:—
   Age 31 yrs. 3 months
   Height 5 ft 11 ins. Eyes Blue
   Complexion Fair Hair Auburn
   Marks/Scars Scar on Forehead
   Trade Group in which employed ~~Truck Driver~~ Issuer
3. Operational Service:—
   (a) Overseas area of service / Embarked from Aust. / Disembarked in Aust.
   ME 5/5 40 27/2 43
   N G 1/8 43 7/3 44
   Borneo 24/3 45 16/10 45
   (b) N.T. (North of Par. 14½° Sth.) or Torres St. Is.:— At From To

Medical Cases Only AMR & O 253A (1) (d) Disability:— Nil
Degree of Disability %
Authenticating Signature Date
For Echelon Use:—

Part B compiled by:— Date 17/9/45 (Officer's signature and rank)

PART C—Compiled by Ech. and Rec.:

4. Non-effective Service:* Tempd Spec Hosp (ME) 23

PART D—Compiled by Ech. & Rec.:

5. Details for Certificate of Discharge No. 228474
   Unit (for discharge purposes) 2/3 Aust Fd Arty
   Served on continuous Full Time War Service in the AIF from 4/11/39 to 26 10 45
   for a Total Effective Period* of 2161 days, which included
   Active Service in Australia† days and 547
   Active Service O/S Australia† 1432 days
   Decorations and Awards‡ during that Service:

War Badge Class and No. RAA A140400 Entered Badge Register 25/10
Discharged from the /and discharge confirmed vide Schedule No. to take effect on and from
Place REDBANK Signature
Date Officer i/c QLD. Ech. and Rec.

6. Details compiled by
   A.A.F. A.101 written by checked by
   Entered Discharge Certificate Register
   Entered "Wills" Register
   A.A.F. A.131 obtained by

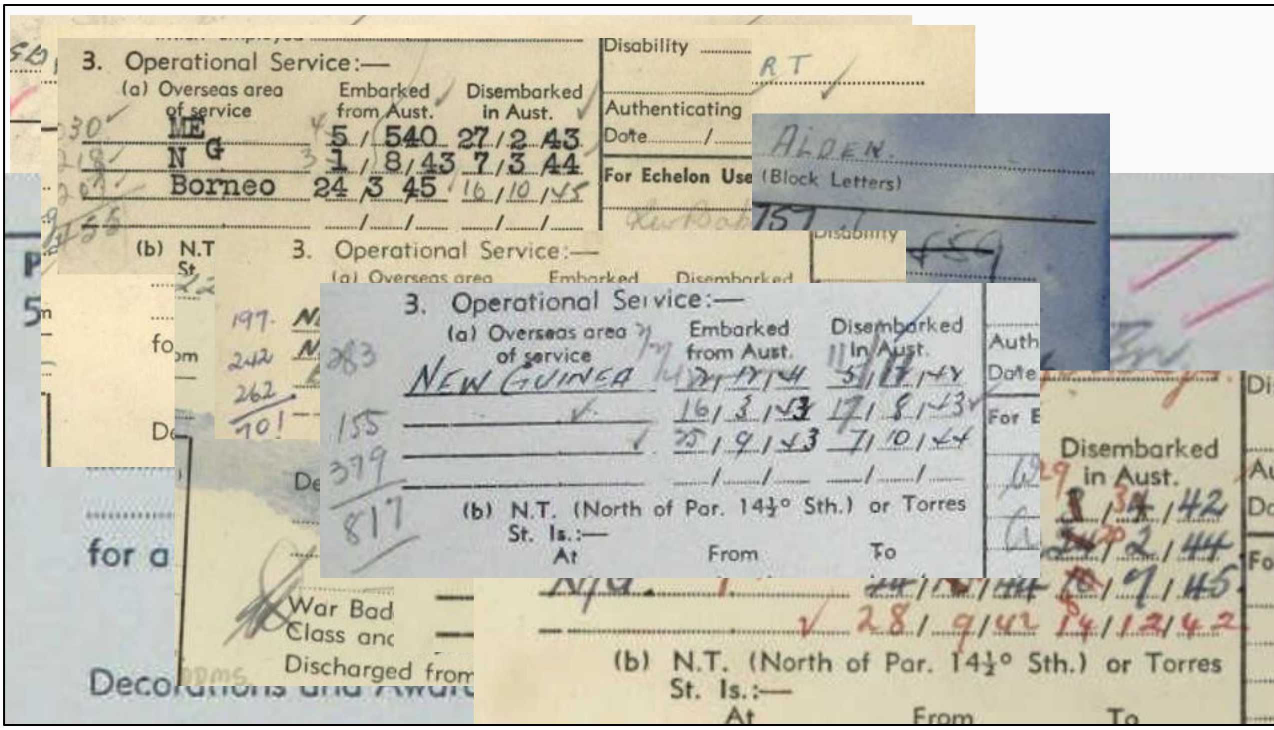
PART E—To be signed by Soldier on discharge:

7. I hereby acknowledge receipt of:
   (a) Certificate of Discharge No. 228474
   (b) Army Form A.131 purporting to contain my Will.
   (c) War Badge No. A140400
   Date Signature of Soldier V.G.A. Keily
   Place REDBANK Signature of Witness

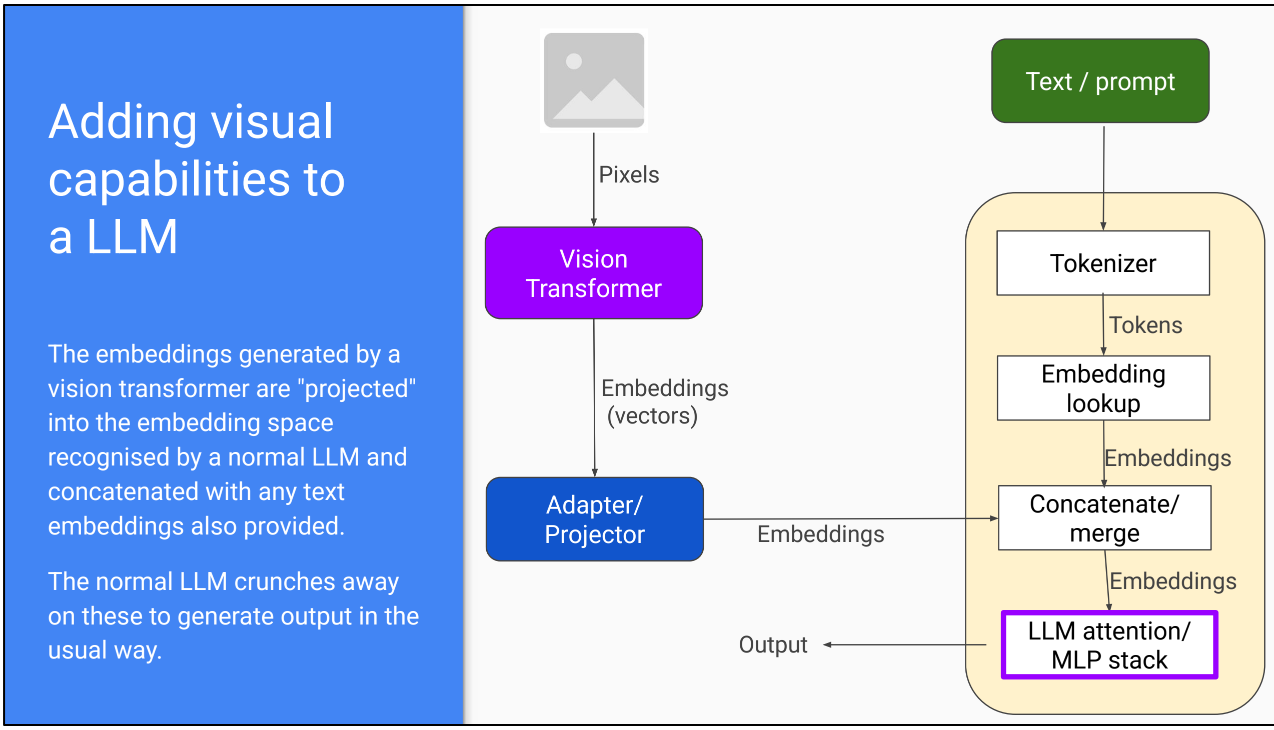
* "Effective Period" means the period of service, less any consecutive 21 days or more for which the soldier was not entitled to pay.
† Australia means the mainland of Australia and Tasmania.
‡ Does not include War Medals.

National Archives of Australia NAA: B883, QX2641

Here's another example - very occasionally information is typed, and as you can see from the top right, it is a bit freestyle - clerks didnt feel compelled to enter data in designated spots



There's also a lot of scribbling and crossing out, and no matter what magnification, lots of content is just hard to interpret…



How can a text LLM handle images?
The stack on the right in brown is a "normal" LLM transformer. It takes in text, and as a first step, tokenises it and then maps those tokens to previously learnt high-dimensional embeddings.
Those embeddings then go through the normal transformer stack, layers upon layers of attention and feed-forward (aka mult level perceptrons) components until finally output tokens are generated.

Vision transformers were initially trained to classify images or label parts of an image, as a cat or a cup or whatever. Some clever person wondered "what would happen if we take the embeddings from a Vision Transformer, and just merge them into the regular embedding stream that a text transformer operates on?"

But the vision transformer was trained on different info, and has "latent space" or "view of the world" completely different from any normal text transformer, so first they had to invent an adapter layer that "projects" the visual embeddings into a high-dimensional embedding space understood by the text transformer. This allows a text transformer to process visual information along with the text.

It is 100% likely that when the visual model was trained, it didnt see many Australian WW2 discharge forms, and if it did, it's brain was swamped by pictures of cats anyway.

So, imagine we want to train a visual model to be able to extract text with specific characteristics from specific places on a specific form.
Our three choices are:

1. Create a new model from scratch - this is too expensive & daunting
2. Take an existing model and fine tune it by training or tweaking or amending all of it's parameters on our data so it gives us the answers we want - this is still expensive as the models are huge.
3. Take an existing mode and tune a small set of extra parameters which, when added to the expression of the existing model's parameters gives us the answers we want. This turns out to be effective and cheap, and is called Parameter Efficient Fine Tuning, and the best known implementation is called LoRA, or Low Rank Adaptation.

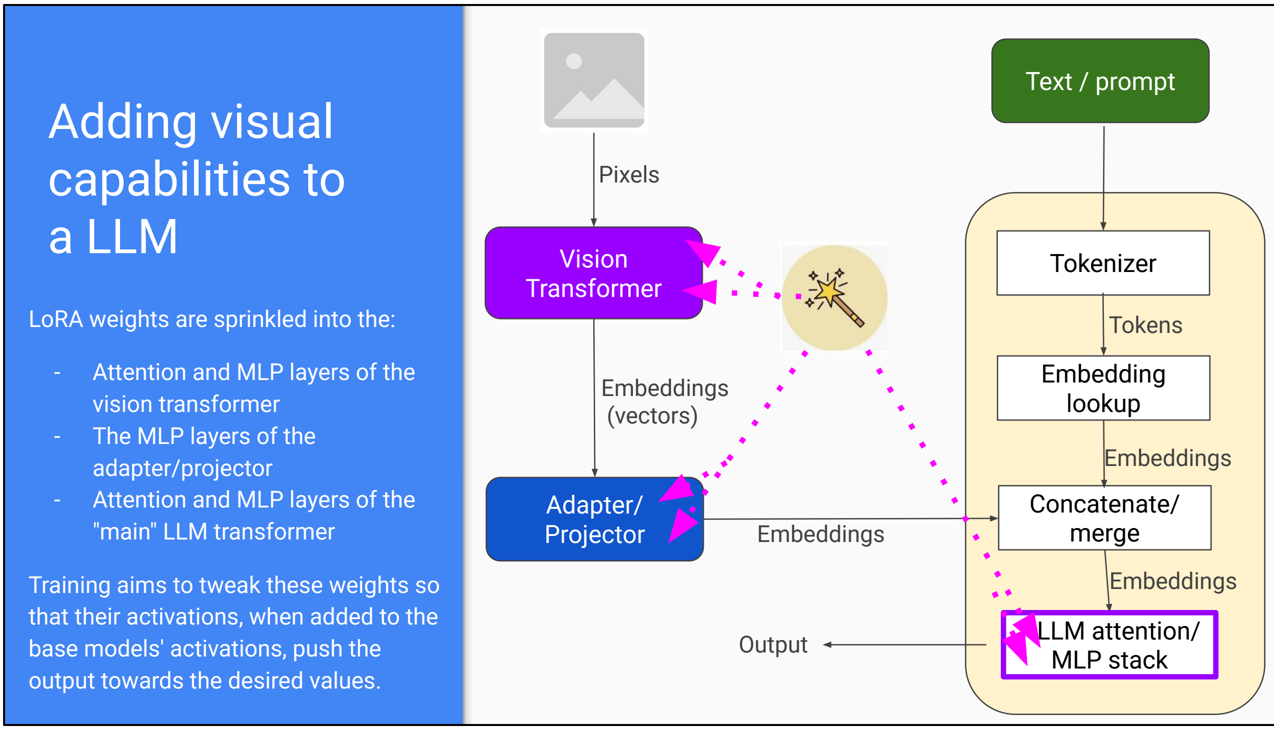
Having loaded a base model, augment it with LoRA

```
model = FastVisionModel.get_peft_model(
  model,
  finetune_vision_layers     = True,
  finetune_language_layers   = True,
  finetune_attention_modules = True,
  finetune_mlp_modules       = True,
  r = 32,  # LoRA rank - larger -> higher Accuracy, but overfit?
  lora_alpha = 32,     # Recommended alpha == r at least
  lora_dropout = 0,
  bias = "none",
  use_rslora = False,  # We support rank stabilized LoRA (?)
  loftq_config = None, # And LoftQ (?)
)
```

Here's a typical configuration of LoRA parameters for a training run.

It defines which parts of the model, vision and/or language, and which components, attention and/or perceptron, will be augmented by LoRA-trained weights.

It also defines how extensive the LoRA weights will be, and hence how much finely tuned modification to the base model can be learnt.



LoRA works by keeping the original model fixed but then adding a tiny % of new parameters/weights to the feed-forward and attention layers throughout the original model.

The fine-tuning I'm discussing here added about 100M new parameters to roughly 8 billion parameter model - around 1%.

LoRA added weights throughout layers in 3 parts the original model - the vision transformer stack, the adapter that maps the vision transformer's embeddings to embeddings the LLM can understand, and the LLM's main transformer stack.

it's not like corrective lenses in glasses stuck to the outside of the model, it is more like deep-brain surgery inserting layers of neurons throughout a brain that alter the outputs of the layers of neurons over which they have been inserted.   Training is comparatively fast because the number of weights being trained is small.

Configure a Supervised Fine Tune Trainer

```
trainer = SFTTrainer(
   model = model,
   tokenizer = tokenizer,
   data_collator =
      UnslothVisionDataCollator(model,
         tokenizer),
   train_dataset =
         converted_train_dataset,
   eval_dataset =
         converted_eval_dataset,
```

```
args = SFTConfig(
    per_device_train_batch_size = 4,
    gradient_accumulation_steps = 4,
    warmup_steps = 5,
    num_train_epochs = 1,
    learning_rate = 1e-4,  # was 2e-4
    logging_steps = 1,
    optim = "adamw_8bit",
    weight_decay = 0.001,
    lr_scheduler_type = "linear",
    seed = 3407,
    fp16_full_eval = True,
    per_device_eval_batch_size = 2,
    eval_accumulation_steps = 4,
    eval_strategy = "steps",
    eval_steps = 8,
    remove_unused_columns = False,
    dataset_text_field = "",
    dataset_kwargs =
      {"skip_prepare_dataset": True},
    max_length = 8192
  ),
)
```

We then configured a Supervised Fine Tune Trainer which will take the training dataset we provide and use it to push the LoRA weights into the direction of causing the overall model to output the results we want.

In our case, we have thousands of images of discharge forms that have already been processed by humans - we know the expected output. This is our training data.

We set aside a small set of images and expected outputs to create an evaluation dataset - the model never learns from these, but is tested on them.

The provided image shows part of a printed WW2 discharge form completed by hand between 1939 and 1948. Focus only on the section under "PART D".

Extract the following information:

1. Below the printed text "Served on continuous Full Time War Service in the" are 1 or 2 rows, each containing three columns:

- Column 1: the service the person served in - either "AIF" or "CMF".
- Column 2: the "from" date.
- Column 3: the "to" date.

Sometimes there will be nothing handwritten in this area, and sometimes only one of two rows will be completed.

Output each completed row as an object in a JSON array named "service" Each service object must contain 3 properties:

- "type": Derived from column 1. Will be "AIF" or "CMF".

- "from": Derived from column 2. Use the corrected date if present (ignore crossed-out dates). Format as dd/mm/yy. Corrections may appear in nearby white space. Dates on the form will not have been written with leading zeros (e.g., "3/5/41" for 3 May 1941). The "to" date will always be after the "from" date in the same row. Be careful not to confuse the pre-printed slashes in " / / " with the numeral "1". Use context to distinguish them. Occasionally a date stamp will have been used to record a date, eg "3 May 1941". In this case, still output the date in dd/mm/yy format.

- "to": Derived from column 3. Same rules as "from".

The final piece of the puzzle is the prompt that tells the model what we want it to do. Here's the Part D prompt. There are separate but similar prompts for the 2 other parts of the form we want to extract.

2. Below the service rows, locate this printed text: "for a total Effective Period* of {totalDays} days, which included". "{totalDays}" will be a handwritten number of days that you must extract as the "totalDays" property.

3. Below this, locate this printed text: "Active Service in Australia† {inAustraliaDays} days, and". "{inAustraliaDays}" will be a handwritten number of days that you must extract as the "inAustraliaDays" property.

4. Below this, locate this printed text: "Active Service O/S Australia† {osAustraliaDays} days, and". "{osAustraliaDays}" will be a handwritten number of days that you must output as the "osAustraliaDays" property.

You only output JSON.

Combine all extracted data into a single JSON object:

{"service": [{"type":..., "from":..., "to":..}...], "totalDays":..., "inAustraliaDays":..,"osAustraliaDays":...}

Example output:

```
{"service":
 [
  {"type":"CMF", "from":"16/04/42",
"to":"12/03/43"},
  {"type":"AIF", "from":"13/03/43", "to":"29/05/46"}
 ],
 "totalDays":1505, "inAustraliaDays":1139,
 "osAustraliaDays":366
}
```

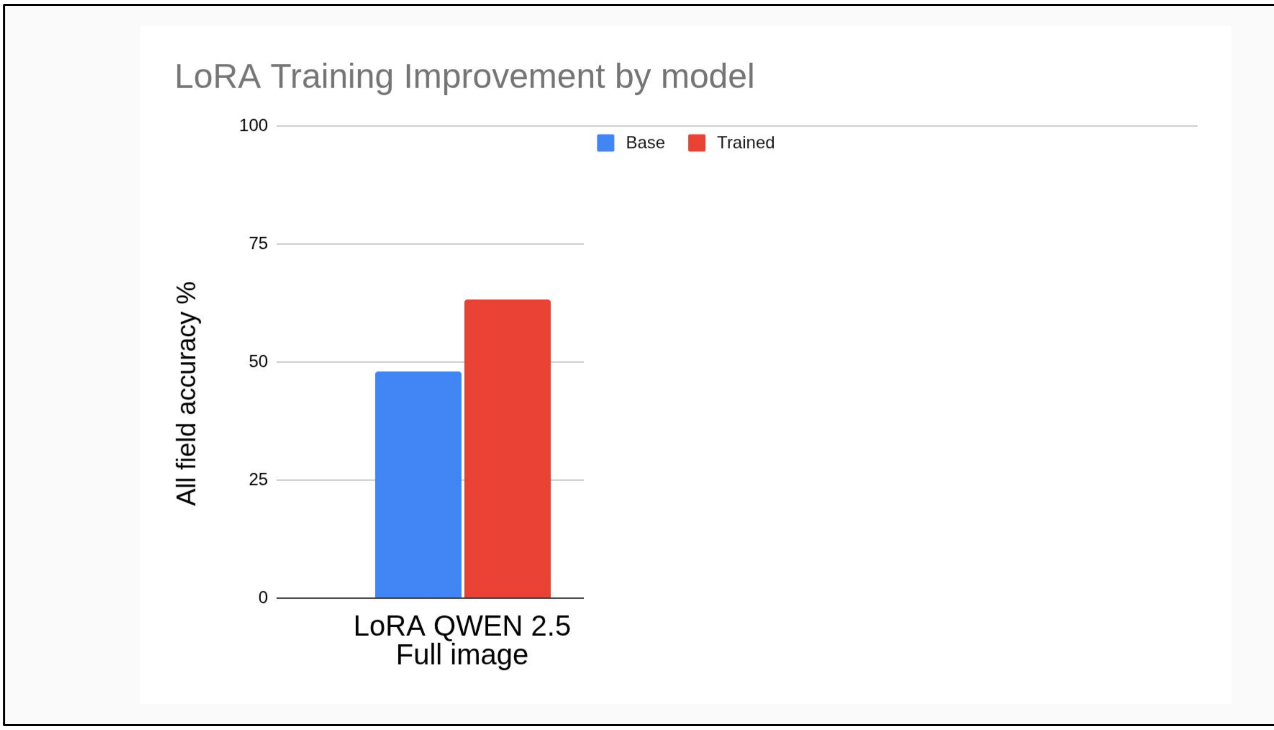
Additional Interpretation Rules:

1. Corrections may be written in nearby white space, not directly over the original, crossed-out entry.
2. Ignore all printed text and any other sections of the form.
3. If "NIL" has been written in any of the "days" fields, output the number 0

Here's more of the same prompt - it is detailed as you can see.

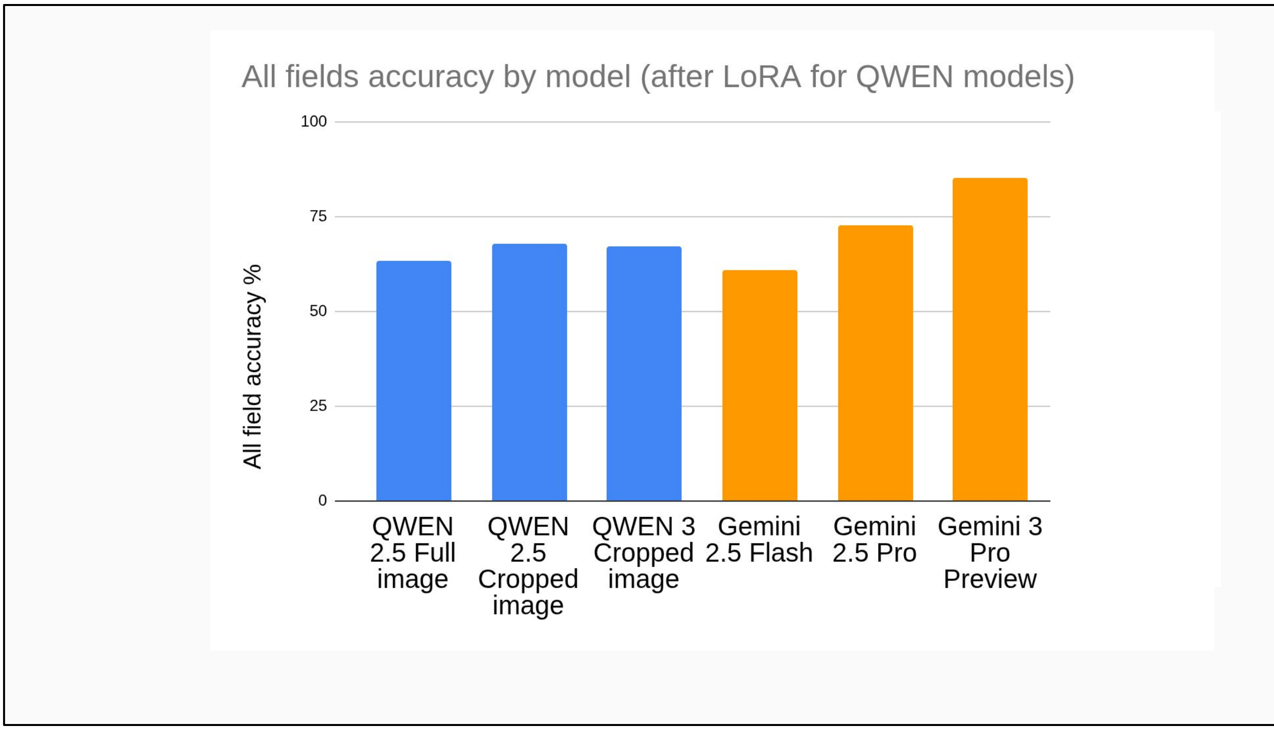
With the training and evaluation datasets prepared (prompts, images, desired answers), we're set to go.

On 2 old Nvidia T4 GPUs it takes about 10 hours to train using about 10,000 training instances divided evenly amongst the 3 prompts.



First try was fine tuning QWEN 2.5 VL 7B, on the full discharge form image . Training increased accuracy on all fields from 48% to 63%.
[ANIMATED]
Next we cropped the image to the area of interest and tried again, hoping removing distractions from irrelevant parts of the form would help. Accuracy increased from 53% to 68%.

Then we tried the new QWEN 3 VL 8B model again on cropped images. Accuracy increased from 45% to 67%.



[ANIMATED]
The evaluation set was also run on the "standard" Gemini 2.5 Flash model, and then on Gemini model, 2.5 Pro which was the best Google model until 2 weeks ago, recently on the latest and greatest, Gemini 3 Pro Preview, which cost about 15 Australian cents to process the 50 evaluation images. Gemini 3 Pro achieved 85% accuracy.

ANIMATED

Humans - 94%

[Updated 14 Dec with improved approaches to QWEN3 fine tuning - see final slide]

Next...

- Train Gemini 3 using Google Vertex AI
    - Training Cost?
    - Performance?
    - Run cost on ~600K discharge forms and other "high value" parts of other forms?
- Track evolution of Deepseek OCR
    - It's radical innovation may realise a leap in trainability and performance

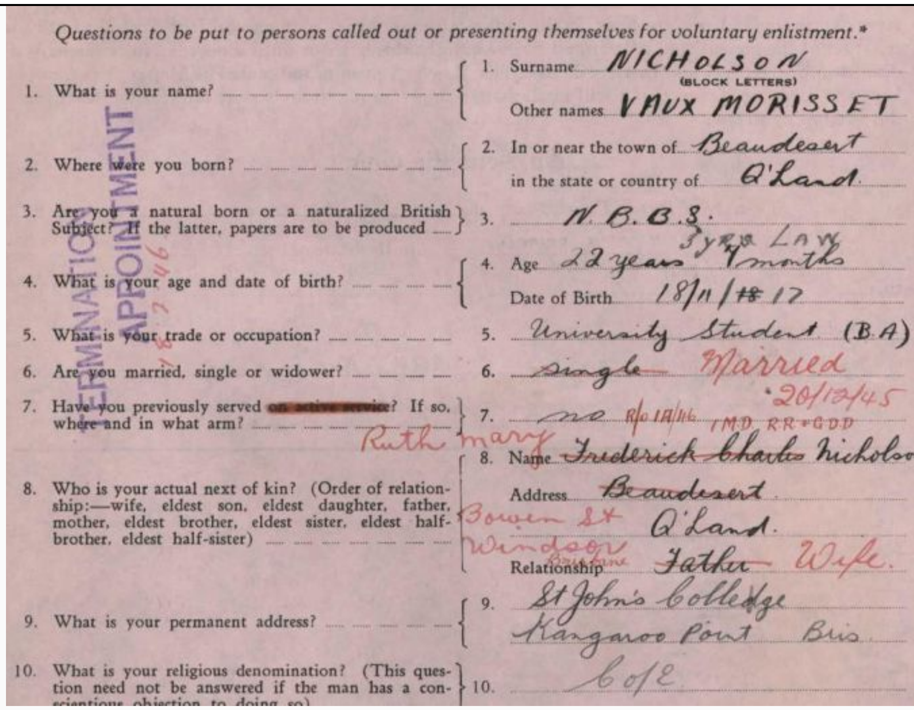
What's next? Fine tuning improved performance of the QWEN models by 15% - 20%. It is plausible Gemini 3 could be fine-tuned with similar effect, taking it from 85% towards or exceeding the 94% human accuracy,
(A small fine-tuning run on Gemini 2.5 Pro using VertexAI improved accuracy across all fields from 73% to 79%)

The cost of automated processing all 600K discharge forms may be about $A6K. Based on previous employment exercises it would cost over $2M to have humans process these forms. The cost of processing all fields of interest on all digitised forms may be about $60K.

However, in practise we know that interpretation of information on a form by humans often requires context of other forms in the file, and deep "domain experience". Also, this data is "rich" so we have a lot to learn as technology improves.

For example [NEXT SLIDE] heres a part of the attestation or enlistment form, showing revisions to marital status, next of kin and address that we want to capture.
Deepseek OCR released just over a month ago is a radical take on how a vision model works, and as it evolves it could leapfrog the current approaches.

The WW2 db will be an incredibly valuable resource for historians, military researchers and the public. Delaying it's availability incurs an annually compounding social cost, yet constructing it with human labour is cost-prohibitive. So it is our responsibility to use the best technology we can to deliver it as soon as possible.

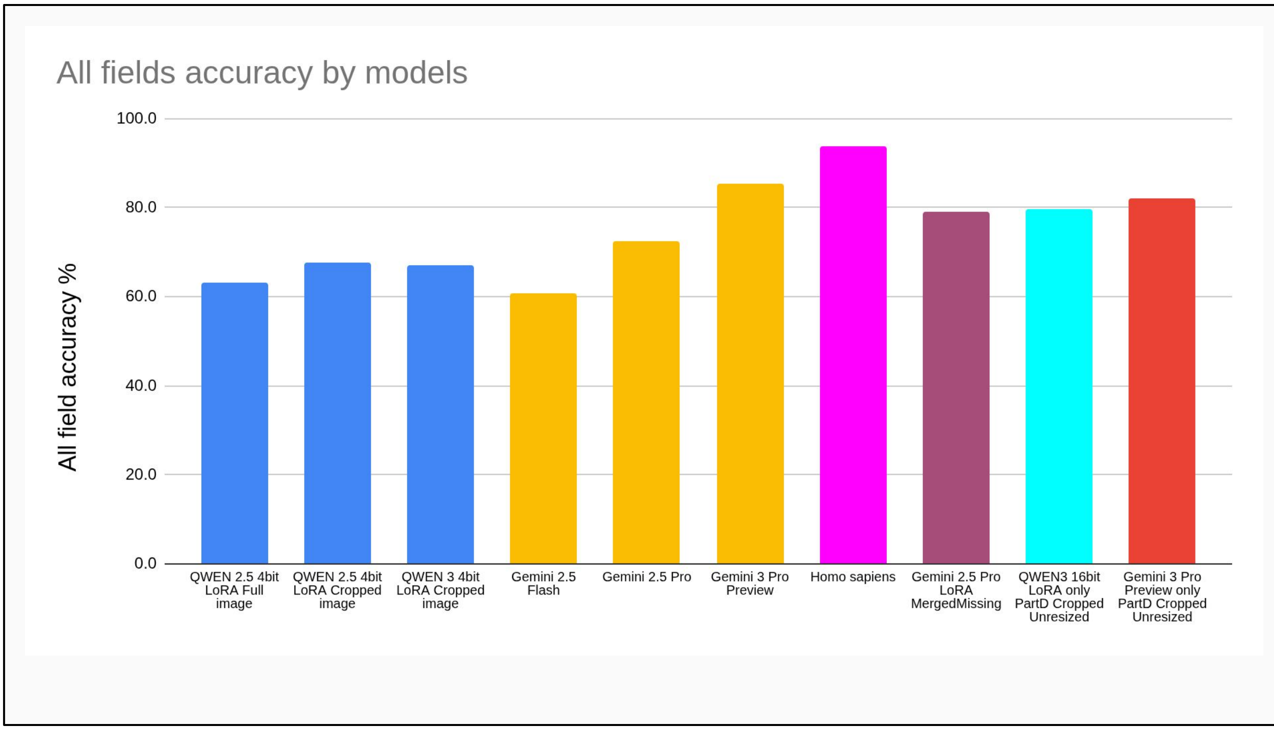


*Questions to be put to persons called out or presenting themselves for voluntary enlistment.**

1. What is your name?	1. Surname NICHOLSON (BLOCK LETTERS) Other names VAUX MORISSET
2. Where were you born?	2. In or near the town of Beaudesert in the state or country of Q'Land
3. Are you a natural born or a naturalized British Subject? If the latter, papers are to be produced	3. N. B. B. S.
4. What is your age and date of birth?	4. Age 22 years 4 months 3yrs LAW Date of Birth 18/11/~~18~~ 17
5. What is your trade or occupation?	5. University Student (B.A)
6. Are you married, single or widower?	6. ~~single~~ Married 20/10/45
7. Have you previously served ~~on active service~~? If so, where and in what arm?	7. no R/o 12/46 IMD RR+GDD
8. Who is your actual next of kin? (Order of relationship:—wife, eldest son, eldest daughter, father, mother, eldest brother, eldest sister, eldest half-brother, eldest half-sister)	8. Name Ruth mary ~~Frederick Charles Nicholson~~ Address ~~Beaudesert~~ Bowen St Windsor Brisbane Q'Land. Relationship ~~Father~~ Wife.
9. What is your permanent address?	9. St John's Colledge Kangaroo Point Bris.
10. What is your religious denomination? (This question need not be answered if the man has a conscientious objection to doing so)	10. C of E

TERMINATION APPOINTMENT 8/2/46

Part of an attestation ("enlistment") form, showing the very common occurrence of a change in marital status and next of kin whilst on service. Some people have multiple such amendments, all of which we capture.



Dec 14 update (not part of AI4LAM presentation)

- Fine tuning Gemini 2.5 Pro gave 79% accuracy on all fields

Subsequently refined fine-tuning on QWEN3 8B VL Instruct

- Only trained on "Part D" part of the form (right hand side: CMF/AIF service dates and the three service category day totals). This is the "hardest" part of the form (most errors in all models). Used native image sizes rather than rescaling to roughly 512 x 512 px crops of this part of the form. For approx 70% of examples, the native image size is larger than 512 x 512 px. Also trained 2 epochs on only this part of the image. So the LoRA adapter is not trained at all in the two other extraction tasks. This approach improved accuracy by QWEN 3 on all fields in Part D from 69% previously to 80%. Gemini 3 Pro Preview given the same (generally larger) clips and the same prompt achieved 82% accuracy on this task (at a cost of $A0.75 for 100 evaluation images).