



Multiklassifiering av svartvita symboler med egna GPU:er

*En jämförande studie av ResNet50 och ResNet18 för
symboligenkänning utan molnbaserad plattform*

- Natalia Timokhova
- Examensarbete 20 yhp
- AI 2025



Innehållsförteckning

1. Introduktion	1
2. Syfte	1
3. Mål	1
4. Metod.....	2
5. Resultat	4
6. Slutsatser och diskussion	8
7. Referenser.....	8

1. Introduktion

Detta examensarbete har utförts i samarbete med TrAI AB, ett företag specialiserat på datorsyn (computer vision) som utvecklar egna produkter inom området. Arbetet bygger vidare på erfarenheter från en praktikperiod (Lärande i Arbete, period 2 – LIA2) mellan januari och maj 2025. Under arbetet upptäcktes att vissa avancerade modeller, såsom YOLO-NAS Medium, hade begränsad möjlighet att tränas på egna datamängder utanför molnplattformar som Roboflow, vilket krävde kommersiella licenser för anpassning. Därför beslutades att minska beroendet av externa plattformar och samtidigt säkerställa en fullt fungerande datorseendemodell.

2. Syfte

Ovannämnda ledde till syftet med detta examensarbete: att utveckla en helautomatiserad pipeline där svartvita symboler annoteras i Roboflow (endast för manuell märkning), exporteras som CSV- och bildfiler, och sedan bearbetas helt på lokala grafikkort med öppna ramverk såsom PyTorch.

Examensarbetet fokuserar specifikt på att jämföra prestandan hos två djupinlärningsarkitekturer - ResNet18 och ResNet50 - för klassificering av symboler i svartvita bilder. Genom att använda öppna modeller och lokala beräkningsresurser kan företag uppnå flera fördelar:

1. Undvika återkommande licenskostnader för molnbaserade tjänster
2. Eliminera molnberoende och uppnå full kontroll över träningsprocessen
3. Möjliggöra lokal träning och återanvändning av modeller
4. Skräddarsy träningsparametrar för optimala resultat på specifika dataset
5. Säkerställa dataintegritet genom att hålla känslig information inom företagets infrastruktur

Detta arbete bidrar till att göra avancerad bildanalys mer tillgänglig och kostnadseffektiv, samtidigt som det ger djupare teknisk kontroll över hela processen.

3. Mål

Det övergripande målet med detta arbete är att utveckla och utvärdera en kostnadseffektiv, lokalt körbar bildklassificeringslösning för symboler i svartvita bilder.

Specifika delmål är:

1. Utveckla en komplett, lokal pipeline för träning och inferens av bildklassificeringsmodeller som eliminerar beroendet av kommersiella molntjänster (t.ex. Roboflow) för inferensfasen och enbart använder öppna, fritt tillgängliga djupinlärningsmodeller.
2. Jämföra prestandan hos ResNet18- respektive ResNet50-arkitekturer för klassificering av svartvita symboler genom systematisk utvärdering av:
 - Tränings- och valideringsnoggrannhet
 - Konvergensmönster och träningshastighet
 - Klassvis prestanda för olika symboltyper
 - Generaliseringsförmåga på verkliga (icke-syntetiska) bilder
 - Resurseffektivitet avseende minnesåtgång, beräkningskraft och energiförbrukning
3. Implementera och använda Weights & Biases (wandb) som verktyg för visualisering och analys av träningsprocessen, för att få djupare insikter i modellernas beteende och prestanda över tid.

4. Metod

Datahantering och förbearbetning

Detta arbete är fokus enbart på datasetet som omfattar fem klasser: 1-polig, dimmer, dubbeltrapp, kron och trapp. Symbolerna annoterades manuellt i Roboflow och exporterades som CSV-filer samt bildsamlingar, organiserade i mapparna train/, valid/ och test/prediction/. Bilderna skalades till 224×224 pixlar och normaliserades med ImageNet-standardvärden (medelvärden [0.485, 0.456, 0.406] och standardavvikelser [0.229, 0.224, 0.225]). För att öka träningsdatavariation augmenterades bilderna med horisontell spegling.

Modellarkitekturer

Två ResNet-varianter jämfördes:

- ResNet18: 18 lager, cirka 11,7 M parametrar
- ResNet50: 50 lager, cirka 25,6 M parametrar

Båda modellerna var förtränade på ImageNet. Det sista fullt anslutna lagret ersattes och anpassades för att klassificera de fem klasserna i Strömbrytare-datasetet. Implementation gjordes i PyTorch, där samtliga förtränade vikter behölls för full finjustering (full fine-tuning).

Träningsprocess

- **Optimerare:** Adam med inlärningstakt $1e-3$
- **Batch-storlek:** 50
- **Förlustfunktion:** Cross-Entropy Loss
- **Epoker:** upp till 40 för att säkerställa konvergens
- **GPU-acceleration:** lokal GPU med optimerad dataladdning (4 arbetstrådar)
- **Modellsparning:** automatiskt sparande av vikter vid bästa valideringsnoggrannhet

Utvärdering

- Noggrannhetsmätningar på tränings-, validerings- och testdata, inklusive klassvis analys
- Visuell inspektion av felklassificerade symboler via en anpassad dashboard
- Jämförelse av konvergensmönster och träningshastighet mellan ResNet18 och ResNet50
- Analys av prestanda på verkliga respektive syntetiska symboler i testdata
- Jämförelse av modellerna före och efter korrigering av felaktiga klassmärkningar i datasetet
- Bedömning av resurseffektivitet med avseende på minnesåtgång, beräkningskraft och energiförbrukning för praktisk implementation i produktionsmiljö

Till skillnad från objekt-detekteringsuppgifter, där mAP (mean Average Precision) används för att utvärdera både lokalisering och klassificering, har detta bildklassificeringsprojekt fokuserat på Accuracy (andel korrekt klassificerade bilder) som huvudmetrik. Accuracy är mest relevant för rena klassificeringsuppgifter där modellen avgör vilken klass en hel, redan isolerad bild tillhör, medan mAP är lämpligare för detektion av objekt inom större bilder.

Visualisering och övervakning

Weights & Biases (wandb) integrerades för att logga och visualisera:

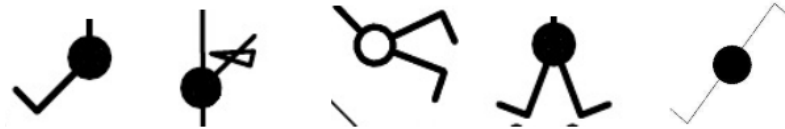
- Tränings- och valideringsförluster
- Tränings- och valideringsnoggrannhet
- Klassvis prestanda för respektive symbolklass

Detta gav djupare insikter i modellernas beteende över tid och underlättade identifiering av eventuella problemområden.

5. Resultat

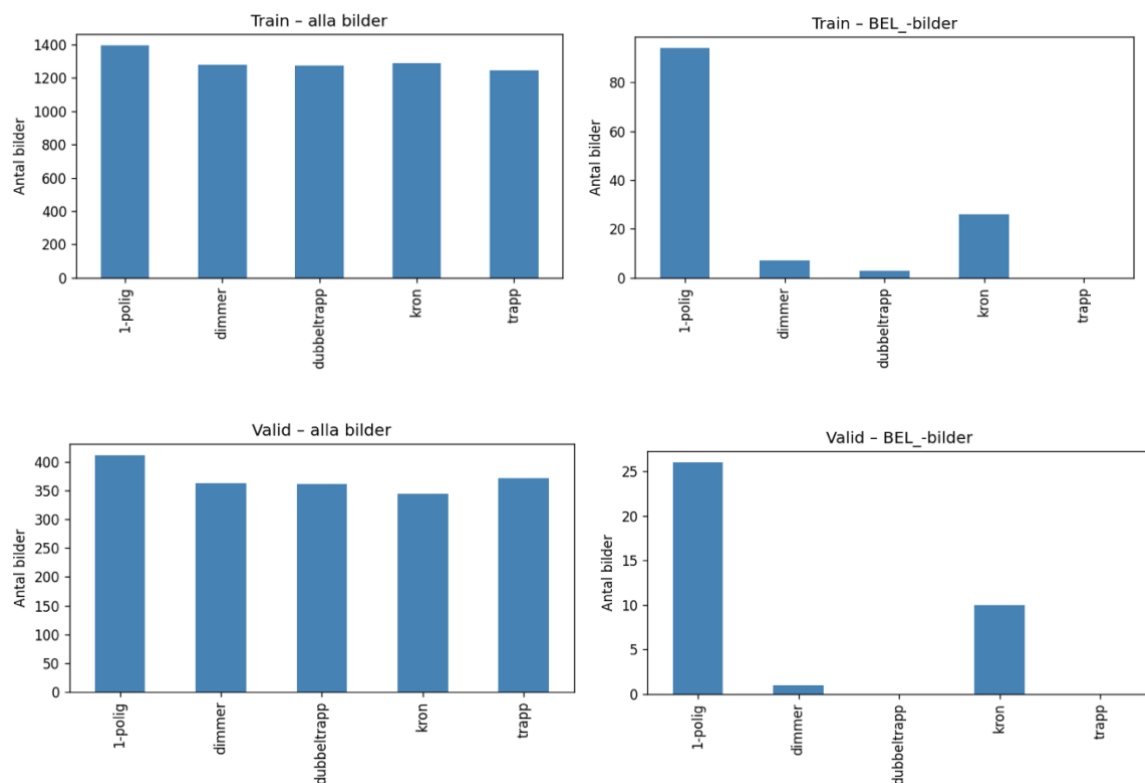
Datautforskning och kvalitetsanalys

Dataset består av 9255 bilder av symboler som man kan se på bilden nedan.

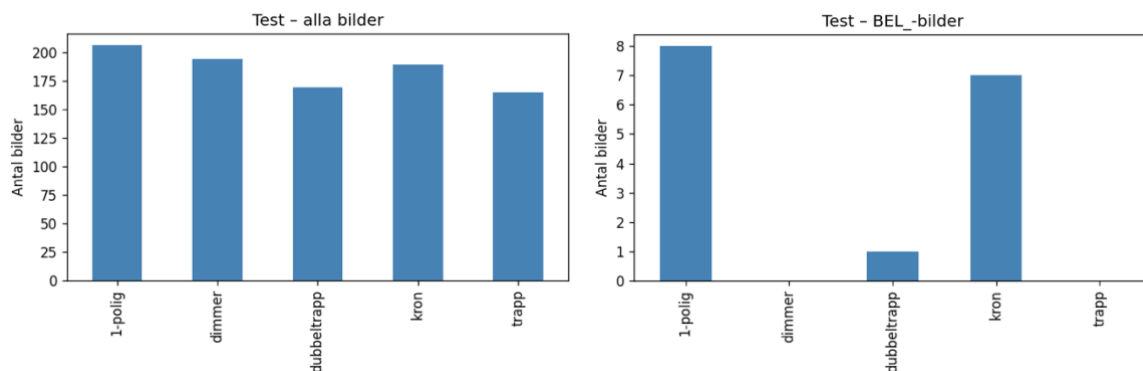


Figur 1. Exempel på symboler i olika klasser i datasetet, från vänster till höger: 1-polig, dimmer, dubbeltrapp, kron, trapp

Datasetet består mestadels av syntetiska bilder (98%) som skapades med syftet att öka antalet bilder i datasetet, samt verkliga bilder som togs från originella bilder (bilder med "BEL_" i namnet, de står till höger på diagrammen nedan). Här kan man se fördelningen av data mellan test-, validerings- och träningsdelarna.



Figur 2. Fördelning av tränings- och validerings dataset, till vänster: syntetiska data, till höger: verkliga data



Figur 3. Fördelning av test datasetet, till vänster: syntetiska data, till höger: verkliga data

Vi konstaterar att datamaterialet är obalanserat, inte bara vad gäller fördelningen mellan verkliga och syntetiska bilder, utan även i antalet förekomster per klass bland de verkliga bilderna. Symbolerna liknar dessutom varandra, särskilt när de är enhetligt vita eller svarta, vilket ledde till att många verkliga symboler felaktigt annoterades under träningen – till exempel annoterades kron-symboler ofta som dubbeltrapp-symboler. Trots det begränsade antalet verkliga symboler är det avgörande att säkerställa hög kvalitet på datasetet som används för modellträning.

Träningsresultat

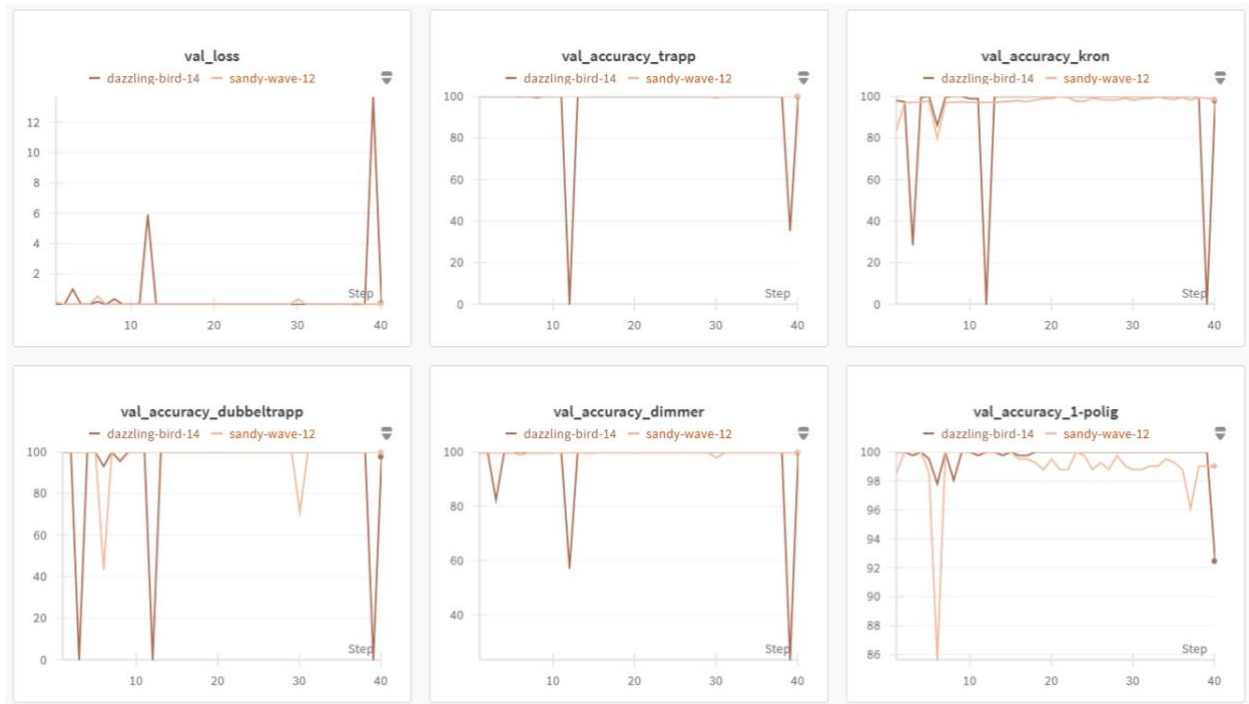
Prestanda på tränings- och valideringsdata:

- ResNet50 uppvisar generellt bättre resultat under både träning och validering.
- Modellen uppnår lägre förlust och högre noggrannhet på såväl tränings- som valideringsdata.
- För varje enskild klass presterar ResNet50 konsekvent bättre eller lika bra som ResNet18.

Tabell 1: Jämförelse av träningsresultat mellan ResNet50 och ResNet18 på det korrigerade datasetet. Värderna representerar prestanda efter fullständig träning.

Metrik	ResNet50	ResNet18
Bästa noggrannhet	99.68%	100.00%
Bästa epoch	21	13
Träningsnoggrannhet	99.74%	99.24%
Träningsförlust	0.0080	0.0247
Valideringsnoggrannhet	99.46%	97.46%
Valideringsförlust	0.0134	0.0748
Klassvis noggrannhet:		
- 1-polig	99.03%	92.46%
- dimmer	99.72%	99.72%
- dubbeltrapp	100.00%	97.79%
- kron	98.55%	97.97%
- trapp	100.00%	100.00%

På de nedanstående diagrammen framgår att ResNet50 (orange) uppvisar lägre förlust än ResNet18 (brun). Det är dock värt att notera att båda modellerna följer liknande mönster med enstaka toppar av hög förlust och perioder med mycket låg noggrannhet. Detta tyder på att datasetet bör granskas avseende felaktiga annotationer och obalanserad klassfördelning innan nästa träningsomgång påbörjas.



Figur 4. Träningresultat av ResNet50 (orange) och ResNet18 (brun)

Tabell 2: Jämförelse av modellernas prestanda på olika datamängder

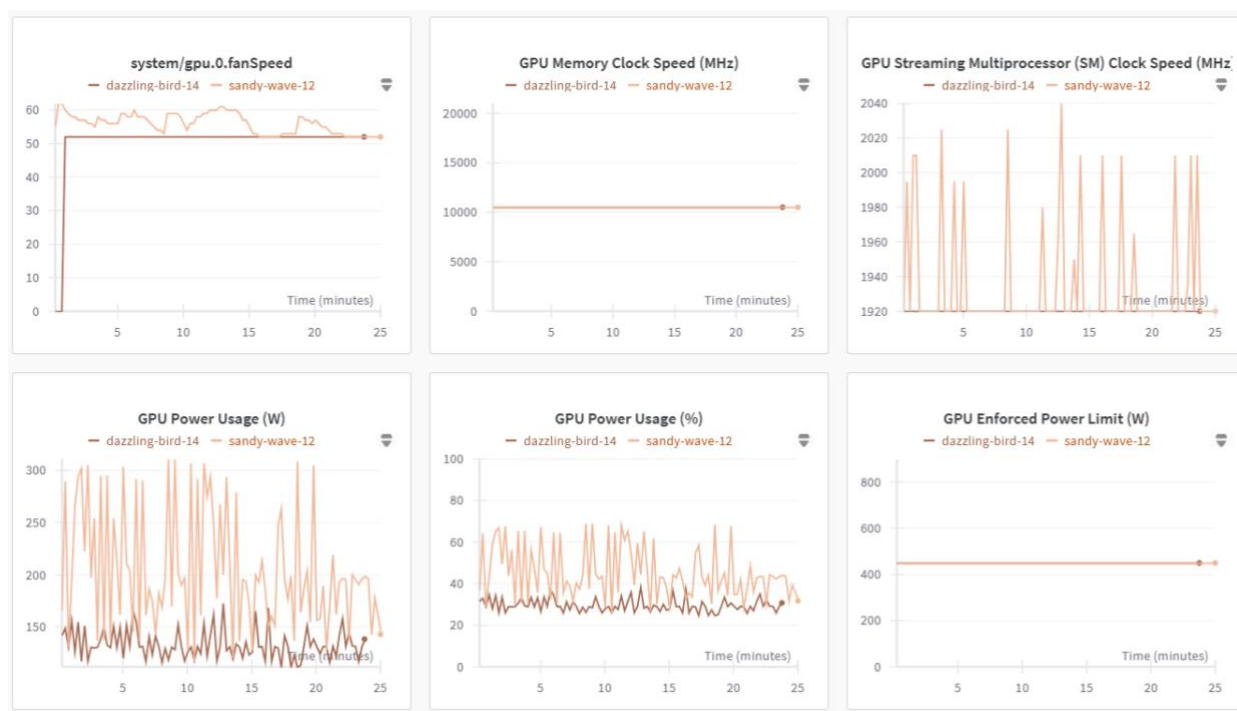
Modell	Valideringsnoggrannhet	Syntetisk+verklig testdata	Endast verklig testdata
ResNet50	99.46%	99.46% (918/923)	68.75% (11/16)
ResNet18	97.46%	99.78% (921/923)	87.50% (14/16)

Prestanda på testdata:

- På hela testdatasetet (främst syntetiskt) presterar båda modellerna exceptionellt bra, med en liten fördel för ResNet18 (99,78 % vs. 99,46 %).
- På den verkliga testdatan är skillnaden tydlig: ResNet18 når 87,50 % noggrannhet, medan ResNet50 når 68,75 %.

Diskrepans mellan syntetiskt dominerande och verkliga testdata:

- Trots högre valideringsnoggrannhet för ResNet50 (99,46 % mot ResNet18:s 97,46 %) presterar den markant sämre på verkliga bilder.
- Detta tyder på att ResNet50 överanpassar sig till syntetiska data och har sämre generaliseringsförmåga på verkliga symboler.



Figur 5. GPU-resursanvändning av ResNet50 (orange) och ResNet18 (brun)

Vid analys av modellernas resursanvändning framkom tydliga skillnader mellan arkitekturerna. ResNet18 slutförde hela träningsprocessen på 23 minuter, medan ResNet50 krävde 25 minuter under identiska förhållanden. Denna tidsskillnad på cirka 8 % kan verka marginell, men grafer över GPU-resursanvändning visar att ResNet50 konsekvent nyttjar avsevärt mer minne och beräkningskraft än sin mindre motsvarighet.

Detta är förväntat med tanke på antalskillnaden i parametrar (11,7 M för ResNet18 jämfört med 25,6 M för ResNet50), men blir särskilt relevant i produktionsmiljöer där flera modeller kan behöva köras parallellt eller där inferens sker i realtid. Den lägre resursåtgången för ResNet18, kombinerat med dess högre noggrannhet på verkliga data, stärker argumentet för att välja den enklare arkitekturen för denna uppgift.

Sammantaget visar jämförelsen att ResNet18 är att föredra framför den mer komplexa ResNet50. Med 87,50 % mot 68,75 % noggrannhet på verkliga data — och liknande prestanda på syntetiska data — bevisar ResNet18 att enklare modeller kan generalisera bättre när träningsdatan domineras av syntetiska exempel. Dess lägre resurskrav gör den dessutom till ett mer praktiskt val i produktionsmiljöer.

6. Slutsatser och diskussion

Projektet har framgångsrikt utvecklat en lokal bildklassificeringslösning för svartvita symboler. ResNet18, trots sin enklare struktur, uppvisade överlägsen generaliseringsförmåga på verkliga data (87,5 % mot 68,75 % för ResNet50). Denna upptäckt utmanar uppfattningen att mer komplexa modeller alltid presterar bättre.

Resursmätningar stärker ytterligare fördelen med den enklare modellen: ResNet18 slutförde träningen på 23 minuter, jämfört med ResNet50:s 25 minuter, samtidigt som den konsekvent använde mindre GPU-minne och beräkningskraft. Denna effektivitet blir särskilt värdefull i produktionsmiljöer där flera modeller kan behöva köras parallellt.

De främsta utmaningarna rörde datasetets kvalitet, med felaktiga annoteringar och obalans mellan syntetiska (98 %) och verkliga (2 %) bilder, vilket särskilt påverkade ResNet50:s generaliseringsförmåga. För framtida förbättringar rekommenderas därför ett mer balanserat dataset och dataaugmenteringstekniker anpassade för att överbrygga gapet mellan syntetiska och verkliga bilder.

Sammanfattningsvis visar projektet att avancerad bildklassificering av svartvita symboler kan implementeras kostnadseffektivt med öppna modeller på lokal infrastruktur, vilket eliminerar beroendet av molnbaserade tjänster samtidigt som resurseffektivitet och hög noggrannhet uppnås.

7. Referenser

1. Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. MIT Press.
2. He, K., Zhang, X., Ren, S. & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. I Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), ss. 770–778.
3. He, K., Zhang, X., Ren, S. & Sun, J. (2016). *Identity Mappings in Deep Residual Networks*. I D. Fleet, T. Pajdla, B. Schiele & T. Tuytelaars (Red.), European Conference on Computer Vision (ECCV) 2016, ss. 630–645. Springer.