



Optical Character Recognition for Myanmar Language

Khin Khant Khant Hlaing

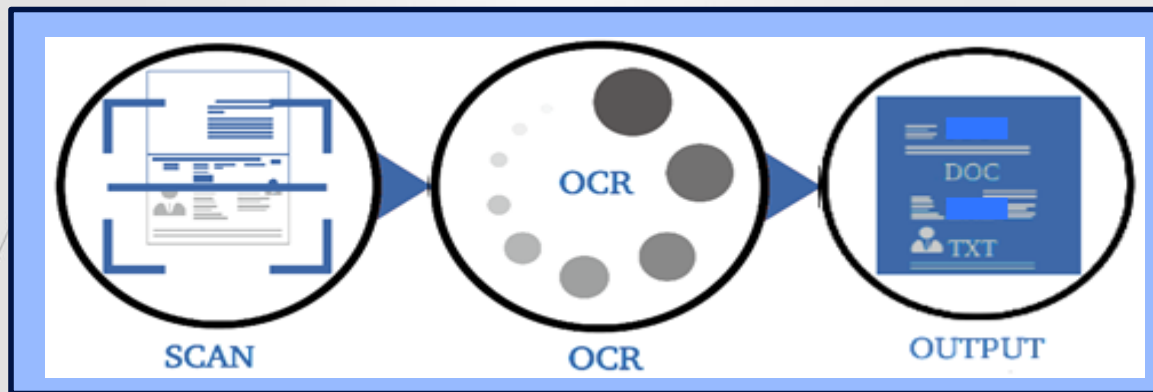
Outline

- **What is OCR ?**
- **Myanmar Script**
- **Implementation of the Myanmar OCR**



What is OCR ?

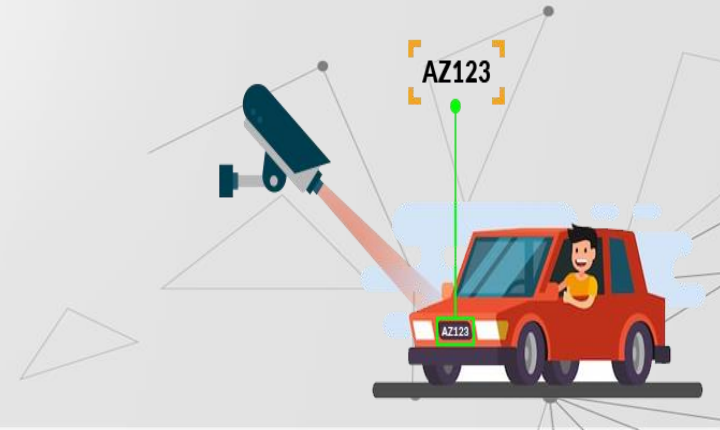
Optical Character Recognition (OCR) is a process of converting an image, such as handwritten files, scanned document or electronic fax, into machine-editable format.



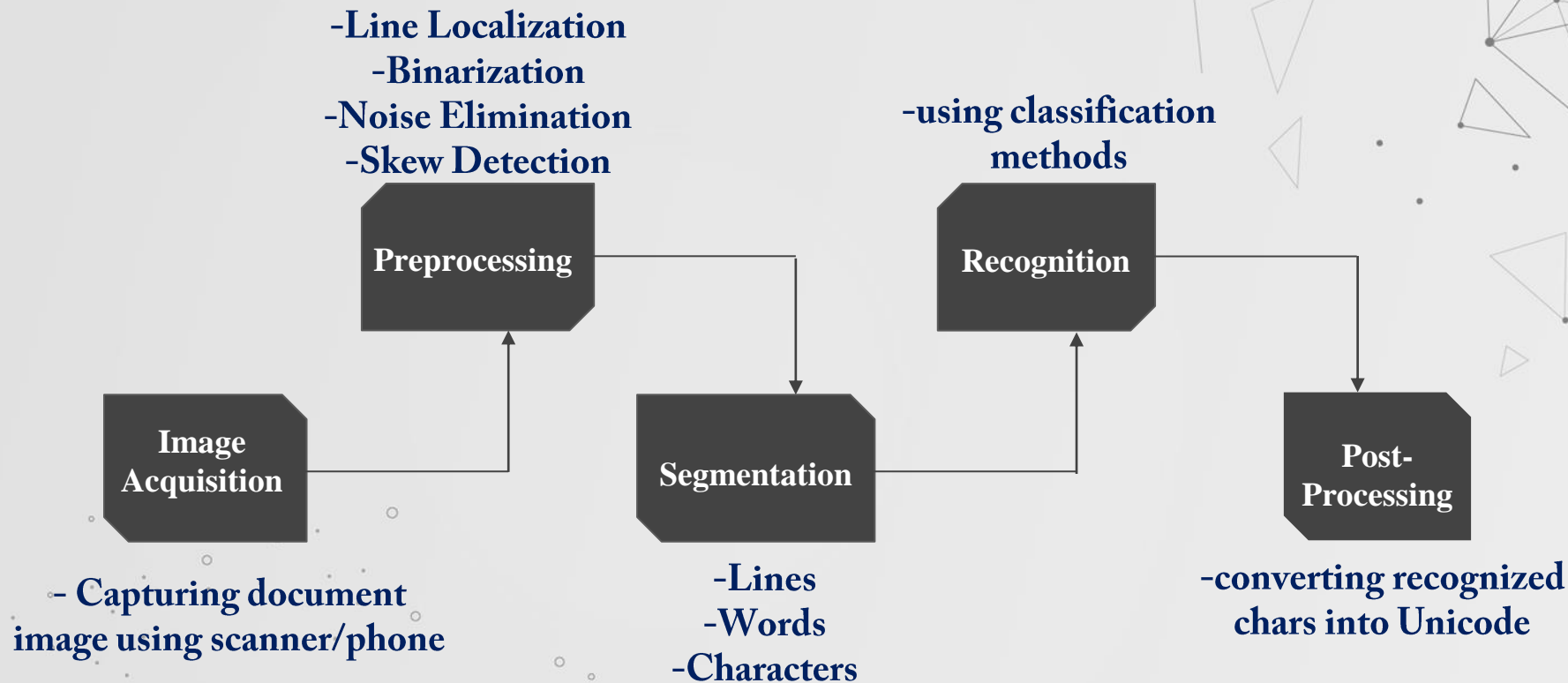


Some usage of the OCR system

- Document Conversion
- Signature Verification and Identification
- Text to Speech
- Online Search



Steps in OCR



An abstract geometric pattern consisting of a network of thin, dark grey lines intersecting at various points. Four of these intersection points are highlighted with solid black dots. The lines radiate from these dots, creating a complex web of triangles and polygons. The overall effect is a minimalist, architectural design.

33 consonants:	ကခဂဃငစဆဇညဋဌဍဎဏတထဒဓနပဖဗဘမယရလဝသဟဋအ
Numbers:	ဝ ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉
Vowels:	a a^{h} e e^{h} i i^{h} o o^{h} u u^{h}
Medial:	w w^{h} y y^{h} r r^{h} l l^{h} v v^{h}
Independent vowels:	က ဤ ဥ ဦ ဧ ဩ ဩ
Symbols and punctuation:	၌ ၍ ၎င်း ၏ ။

- The Myanmar script developed from the Mon script and more than (32) million people use Myanmar script.
- It is written from left to right and requires no spaces between words.

Figure. A set of Myanmar Alphabet

Implementation of the Myanmar OCR

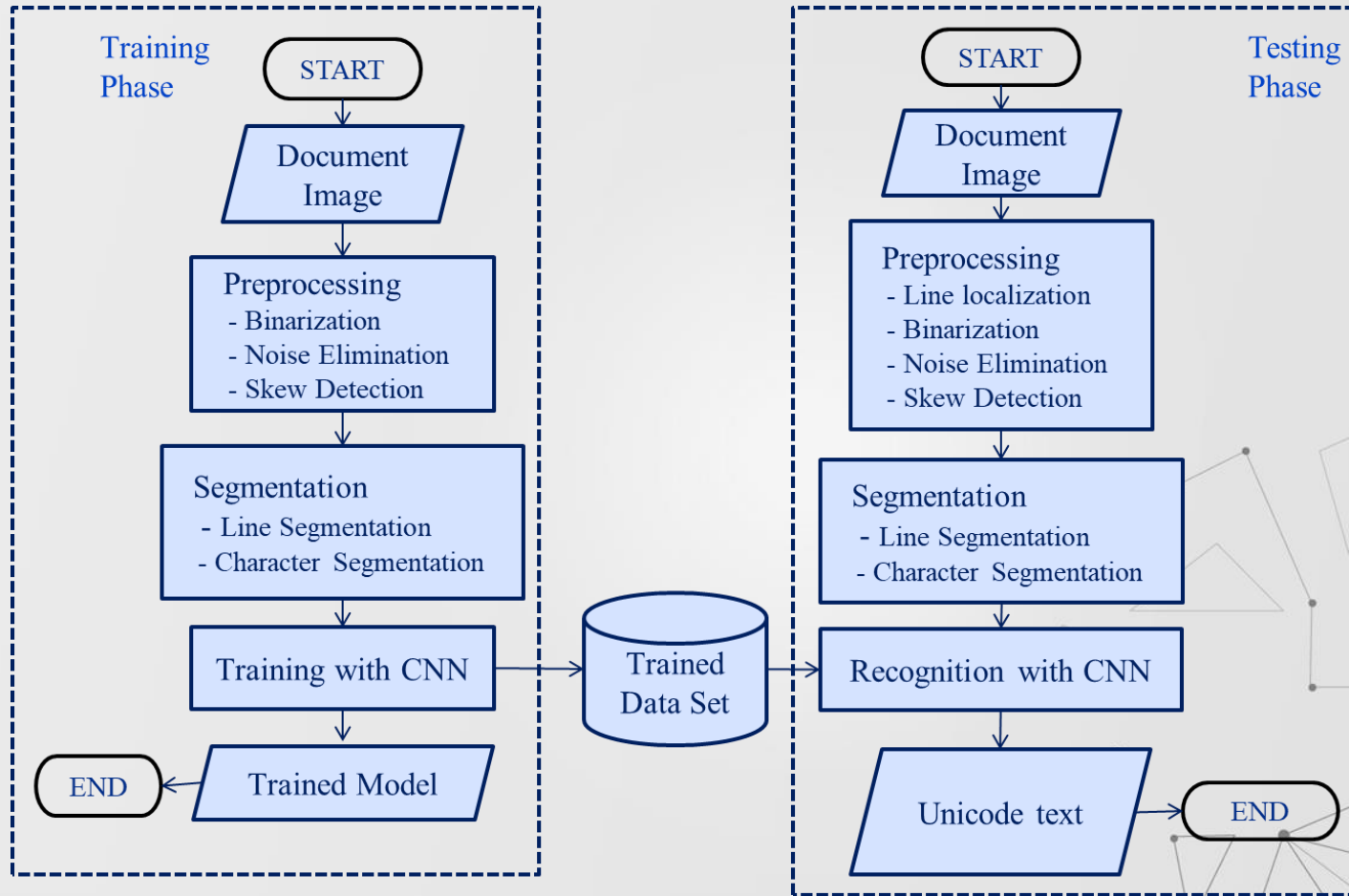


Image Acquisition

The document image, in tiff, jpg, bmp, and png format is captured using a IR-3245N scanner and phone.



ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်
နိုင်ငံတော်သမ္မတ ဦးဝင်းမြင့် ထံမှ
အလုပ်သမားနေ့အခမ်းအနားသို့ ပေးပို့သော
သဝဏ်လွှာ

(၂၀၁၉ ခုနှစ်၊ မေလ ၁ ရက်)

ချစ်ကြည်လေးစားအပ်ပါသော အလုပ်သမားကြီးများခင်ဗျား-

ယနေ့ကျရောက်သည့် အလုပ်သမားနေ့အခါသမယမှာ ပြည်ပနိုင်ငံအသီးသီး၌ ရောက်ရှိနေကြသည့် မြန်မာရွှေ့ပြောင်းအလုပ်သမားများအပါအဝင် မြန်မာအလုပ်သမားတုတစ်ရပ်လုံး ကိုယ်၏ကျန်းမာခြင်း၊ စိတ်၏ချမ်းသာခြင်း ပြည့်ဝပြီး ကောင်းကျိုးမင်္ဂလာ အဖြာဖြာနှင့် ပြည့်စုံကြပါစေ၊ လုပ်ငန်းခွင်အေးချမ်းသာယာပြီး လူမှုစီးပွားဘဝများ ဖွံ့ဖြိုးတိုးတက်ကြပါစေလို့ ဆုတောင်းမေတ္တာပို့သရင်း ဦးစွာနှုတ်ခွန်းဆက်သအပ်ပါသည်။

ကမ္ဘာတစ်ဝန်းရှိ အလုပ်သမားများကို အလေးထားဂုဏ်ပြုသည့်အနေဖြင့် ပေါ်ပေါက်ခဲ့သော အလုပ်သမားနေ့သည် ခေတ်အဆက်ဆက် ရှည်လျားသည့် သမိုင်းကြောင်းနှင့် အတူ ကျော်ဖြတ်ခဲ့ရခြင်းဖြစ်ပါသည်။ စက်မှုတော်လှန်ရေးအဆက်ဆက်တွင် အလုပ်သမားများအနေနှင့် ရသင့်ရထိုက်သည့် အခွင့်အရေးများ၊ လိုလားချက်များကို အခက်အခဲမျိုးစုံကြားမှ တောင်းဆိုရယူနိုင်ခဲ့ကြပါသည်။ လက်ရှိ ကာလတွင် မြန်မာနိုင်ငံအပါအဝင် ကမ္ဘာတစ်ဝန်းမှ အလုပ်သမားများ၏ ရသင့်ရထိုက်သည့် အခွင့်အရေးများသည် ယခင်ကထက် ပိုမိုတိုးတက်ရရှိနေကြောင်း တွေ့မြင်နိုင်ပါသည်။ ယခုကဲ့သို့ တိုးတက်မှုများ ဖြစ်ပေါ်လာရန် စိတ်ထားကောင်းနှင့် ဦးဆောင်ဦးရွက်ပြုခဲ့သည့် အလုပ်သမားများကို လေးစားအသိအမှတ်ပြုအပ်ပါသည်။ အလုပ်သမားများ ရသင့်ရထိုက်သည့် အခွင့်အရေးများကို ဥပဒေနှင့်အညီ ရရှိခံစားနိုင်ရန် နိုင်ငံတော်အစိုးရအနေဖြင့် အလေးထားဆောင်ရွက်ပေးလျက်ရှိပါသည်။

Hough transform is performed on the document image for skew detection.

[illegible]



ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်
နိုင်ငံတော်သမ္မတ ဦးဝင်းမြင့် ဟု
အလုပ်သမားနေ့အခမ်းအနားသို့ ပေးပို့သော
သတင်းလွှာ
(၂၀၁၅ ခုနှစ် မေလ ၁ ရက်)

ခုတ်ကြည့်ကဝေအစခေါ်ပါသော အလုပ်သမား ပြဿနာတစ်ရပ်

ကနေ့ကျရောက်သည့် အလုပ်သမားနေ့အခမ်းအနားမှာ ပြည်တစ်ရပ်လုံးလုံး
ခရောင်မြောက်သည့် မြန်မာ့ရွှေပြောင်အလုပ်သမားများအပေါ်ကပ် စွဲနေကလွန်သောအခါ
တစ်ခုလုံး ကိုယ်စားပြုနေပါမိလိမ့်။ စိတ်ဓါတ်ချမ်းသာခြင်း ပြည့်ဝပြီး တောင်းချက်သစ်
သဖြင့်ဖြင့် ပြည့်စုံကြပါစေ လုပ်ငန်းခွင်အေးချမ်းသာယာပြီး လူမှုစီးပွားကလေး ဖွံ့ဖြိုး
တိုးတက်ကြပါစေလို့ ဆုတောင်းဆုတောင်းလျက် ဦးစွာ နှုတ်ခွင့်တောင်းခံပါသည်။

ယူဆသမိမ့်၍ အကုန်လုံးအလုပ်ကို ဘဝတစ်ခတ်တည်းပြုကုန်ခဲ့သဖြင့် ပေါက်ကွဲ
ခဲ့သော အလုပ်သမားအနေဖြင့် တောင်းဆိုတောင်းဆို ဂုဏ်ကျသည့် သနိဋ္ဌိကြောင်းနှင့်
အလုပ်သမားပြုခြင်းနှင့် ဖြစ်လာသည့် ခက်ခဲမှုအခက်များအပေါ် သတိမထားမိဘဲ အလုပ်သမား
များအနေဖြင့် အလုပ်သမားပြုမှု အခွင့်အရေးများ ကိုလေးစားမှုရှိမှုကို အသက်အသံ
ဖြင့်ပြောရာ တောင်းဆိုမှုနှင့် ဖြစ်ပါသည်။ ပေးသည့် ဝယ်ယူမှုမြန်နှုန်းနှင့် အသက်အသံ
ကလေးတစ်ယောက် အလုပ်သမားများ၏ ရသင့်ရင့်ကဲကဲနှင့် အခွင့်အရေးများအသံ
သတိထားမိမိ ဝိန်းကွဲသော နှုတ်ခွင့်ကြောင့် အခွင့်အရေးတစ်ခုလုံးကို အသက်အသံ
ဖြင့်ပြောရာ တောင်းဆိုမှုနှင့် ဖြစ်ပါသည်။ ဦးစွာ နှုတ်ခွင့်တောင်းခံပြီး ကလေးတစ်ယောက်
လေးတစ်ယောက်အနေဖြင့် ပြောပါသည်။ အလုပ်သမားများ နည်းစနစ်တစ်ခုနှင့် အခွင့်အရေး
များကို ဥပဒေမရှိမည့် ချိန်ခွဲသမားနှင့် စိတ်ဝင်တစားအစိုးရအနေဖြင့် အလေးထား
သောလုပ်ငန်းအခမ်းအနားတစ်ခု



ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်
နိုင်ငံတော်သမ္မတ ဦးဝင်းမြင့် ဟု
အလုပ်သမားနေ့အခမ်းအနားသို့ ပေးပို့သော
သတင်းလွှာ
(၂၀၁၅ ခုနှစ် မေလ ၁ ရက်)

ခုတ်ကြည့်ကဝေအစခေါ်ပါသော အလုပ်သမား ပြဿနာတစ်ရပ်

ကနေ့ကျရောက်သည့် အလုပ်သမားနေ့အခမ်းအနားမှာ ပြည်တစ်ရပ်လုံးလုံး
ခရောင်မြောက်သည့် မြန်မာ့ရွှေပြောင်အလုပ်သမားများအပေါ်ကပ် စွဲနေကလွန်သောအခါ
တစ်ခုလုံး ကိုယ်စားပြုနေပါမိလိမ့်။ စိတ်ဓါတ်ချမ်းသာခြင်း ပြည့်ဝပြီး တောင်းချက်သစ်
သဖြင့်ဖြင့် ပြည့်စုံကြပါစေ လုပ်ငန်းခွင်အေးချမ်းသာယာပြီး လူမှုစီးပွားကလေး ဖွံ့ဖြိုး
တိုးတက်ကြပါစေလို့ ဆုတောင်းဆုတောင်းလျက် ဦးစွာ နှုတ်ခွင့်တောင်းခံပါသည်။

ယူဆသမိမ့်၍ အကုန်လုံးအလုပ်ကို ဘဝတစ်ခတ်တည်းပြုကုန်ခဲ့သဖြင့် ပေါက်ကွဲ
ခဲ့သော အလုပ်သမားအနေဖြင့် တောင်းဆိုတောင်းဆို ဂုဏ်ကျသည့် သနိဋ္ဌိကြောင်းနှင့်
အလုပ်သမားပြုခြင်းနှင့် ဖြစ်လာသည့် ခက်ခဲမှုအခက်များအပေါ် သတိမထားမိဘဲ အလုပ်သမား
များအနေဖြင့် အလုပ်သမားပြုမှု အခွင့်အရေးများ ကိုလေးစားမှုရှိမှုကို အသက်အသံ
ဖြင့်ပြောရာ တောင်းဆိုမှုနှင့် ဖြစ်ပါသည်။ ပေးသည့် ဝယ်ယူမှုမြန်နှုန်းနှင့် အသက်အသံ
ကလေးတစ်ယောက် အလုပ်သမားများ၏ ရသင့်ရင့်ကဲကဲနှင့် အခွင့်အရေးများအသံ
သတိထားမိမိ ဝိန်းကွဲသော နှုတ်ခွင့်ကြောင့် အခွင့်အရေးတစ်ခုလုံးကို အသက်အသံ
ဖြင့်ပြောရာ တောင်းဆိုမှုနှင့် ဖြစ်ပါသည်။ ဦးစွာ နှုတ်ခွင့်တောင်းခံပြီး ကလေးတစ်ယောက်
လေးတစ်ယောက်အနေဖြင့် ပြောပါသည်။ အလုပ်သမားများ နည်းစနစ်တစ်ခုနှင့် အခွင့်အရေး
များကို ဥပဒေမရှိမည့် ချိန်ခွဲသမားနှင့် စိတ်ဝင်တစားအစိုးရအနေဖြင့် အလေးထား
သောလုပ်ငန်းအခမ်းအနားတစ်ခု



ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်
နိုင်ငံတော်သမ္မတ ဦးဝင်းမြင့် ထံမှ
အလုပ်သမားနေ့အခမ်းအနားသို့ ပေးပို့သော
သတင်းလွှာ
(၂၀၁၉ ခုနှစ်၊ မေလ ၁ ရက်)

ချစ်ကြည်လေးစားအပ်ပါသော အလုပ်သမားကြီးများခင်ဗျား-
ယနေ့ကျရောက်သည့် အလုပ်သမားနေ့အခမ်းအနားမှာ ပြည်ပနိုင်ငံအသီးသီးသို့
ရောက်ရှိနေကြသည့် မြန်မာရွှေ့ပြောင်းအလုပ်သမားများအပါအဝင် မြန်မာအလုပ်သမားထု
တစ်ရပ်လုံး ကိုယ်စိတ်ကုန်းမာခြင်း၊ စိတ်၏ချမ်းသာခြင်း၊ ပြည့်ဝပြီး ကောင်းကျိုးမက်လာ
အပြောနှင့် ပြည့်စုံကြပါစေ၊ လုပ်ငန်းခွင်အချမ်းသာယာပြီး လူမှုစီးပွားဘဝများ စွဲမြဲ
တိုးတက်ကြပါစေလို့ ဆုတောင်းပေးတာပါ။ သူတို့အား အလုပ်သမားများကို အလေးထားစုတ်ပြုသည့်အနေဖြင့် ပေါ်ပေါက်
ခဲ့သော အလုပ်သမားနေ့သည် ခေတ်အဆက်ဆက် ဖြည့်လျားသည့် သမိုင်းကြောင်းနှင့်
အတူ ကော်ဗြစ်ခုံရံခြင်းဖြစ်ပါသည်။ စက်မှုတော်လှန်ရေးအဆက်ဆက်တွင် အလုပ်သမား
များအနေနှင့် ရသင့်ရထိုက်သည့် အခွင့်အရေးများ၊ လိုလားချက်များကို အခက်အခဲ
မျိုးစုံကြားမှ တောင်းဆိုရယူနိုင်ခဲ့ကြပါသည်။ လက်ရှိ ကာလတွင် မြန်မာနိုင်ငံအပါအဝင်
ကမ္ဘာတစ်ဝန်းမှ အလုပ်သမားများ၏ ရသင့်ရထိုက်သည့် အခွင့်အရေးများသည်
ယခင်ကထက် ပိုမိုတိုးတက်ကျန်နေသကဲ့သို့ တွေ့မြင်နိုင်ပါသည်။ ယခုကဲ့သို့ တိုးတက်မှုများ
ဖြစ်ပေါ်လာရန် စိတ်ထားကောင်းနှင့် ဦးဆောင်ဦးရွက်ပြုခဲ့သည့် အလုပ်သမားများကို
လေးစားအသိအမှတ်ပြုအပ်ပါသည်။ အလုပ်သမားများ ရသင့်ရထိုက်သည့် အခွင့်အရေး
များကို ဥပဒေနှင့်အညီ ရရှိခံစားနိုင်ရန် နိုင်ငံတော်အစိုးရအနေဖြင့် အလေးထား
ဆောင်ရွက်ပေးလျက်ရှိပါသည်။

ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်
နိုင်ငံတော်သမ္မတ ဦးဝင်းမြင့် ထံမှ
အလုပ်သမားနေ့အခမ်းအနားသို့ ပေးပို့သော
သတင်းလွှာ
(၂၀၁၉ ခုနှစ်၊ မေလ ၁ ရက်)

ချစ်ကြည်လေးစားအပ်ပါသော အလုပ်သမားကြီးများခင်ဗျား-
ယနေ့ကျရောက်သည့် အလုပ်သမားနေ့အခမ်းအနားမှာ ပြည်ပနိုင်ငံအသီးသီးသို့
ရောက်ရှိနေကြသည့် မြန်မာရွှေ့ပြောင်းအလုပ်သမားများအပါအဝင် မြန်မာအလုပ်သမားထု
တစ်ရပ်လုံး ကိုယ်စိတ်ကုန်းမာခြင်း၊ စိတ်၏ချမ်းသာခြင်း၊ ပြည့်ဝပြီး ကောင်းကျိုးမက်လာ
အပြောနှင့် ပြည့်စုံကြပါစေ၊ လုပ်ငန်းခွင်အချမ်းသာယာပြီး လူမှုစီးပွားဘဝများ စွဲမြဲ
တိုးတက်ကြပါစေလို့ ဆုတောင်းပေးတာပါ။ သူတို့အား အလုပ်သမားများကို အလေးထားစုတ်ပြုသည့်အနေဖြင့် ပေါ်ပေါက်
ခဲ့သော အလုပ်သမားနေ့သည် ခေတ်အဆက်ဆက် ဖြည့်လျားသည့် သမိုင်းကြောင်းနှင့်
အတူ ကော်ဗြစ်ခုံရံခြင်းဖြစ်ပါသည်။ စက်မှုတော်လှန်ရေးအဆက်ဆက်တွင် အလုပ်သမား
များအနေနှင့် ရသင့်ရထိုက်သည့် အခွင့်အရေးများ၊ လိုလားချက်များကို အခက်အခဲ
မျိုးစုံကြားမှ တောင်းဆိုရယူနိုင်ခဲ့ကြပါသည်။ လက်ရှိ ကာလတွင် မြန်မာနိုင်ငံအပါအဝင်
ကမ္ဘာတစ်ဝန်းမှ အလုပ်သမားများ၏ ရသင့်ရထိုက်သည့် အခွင့်အရေးများသည်
ယခင်ကထက် ပိုမိုတိုးတက်ကျန်နေသကဲ့သို့ တွေ့မြင်နိုင်ပါသည်။ ယခုကဲ့သို့ တိုးတက်မှုများ
ဖြစ်ပေါ်လာရန် စိတ်ထားကောင်းနှင့် ဦးဆောင်ဦးရွက်ပြုခဲ့သည့် အလုပ်သမားများကို
လေးစားအသိအမှတ်ပြုအပ်ပါသည်။ အလုပ်သမားများ ရသင့်ရထိုက်သည့် အခွင့်အရေး
များကို ဥပဒေနှင့်အညီ ရရှိခံစားနိုင်ရန် နိုင်ငံတော်အစိုးရအနေဖြင့် အလေးထား
ဆောင်ရွက်ပေးလျက်ရှိပါသည်။

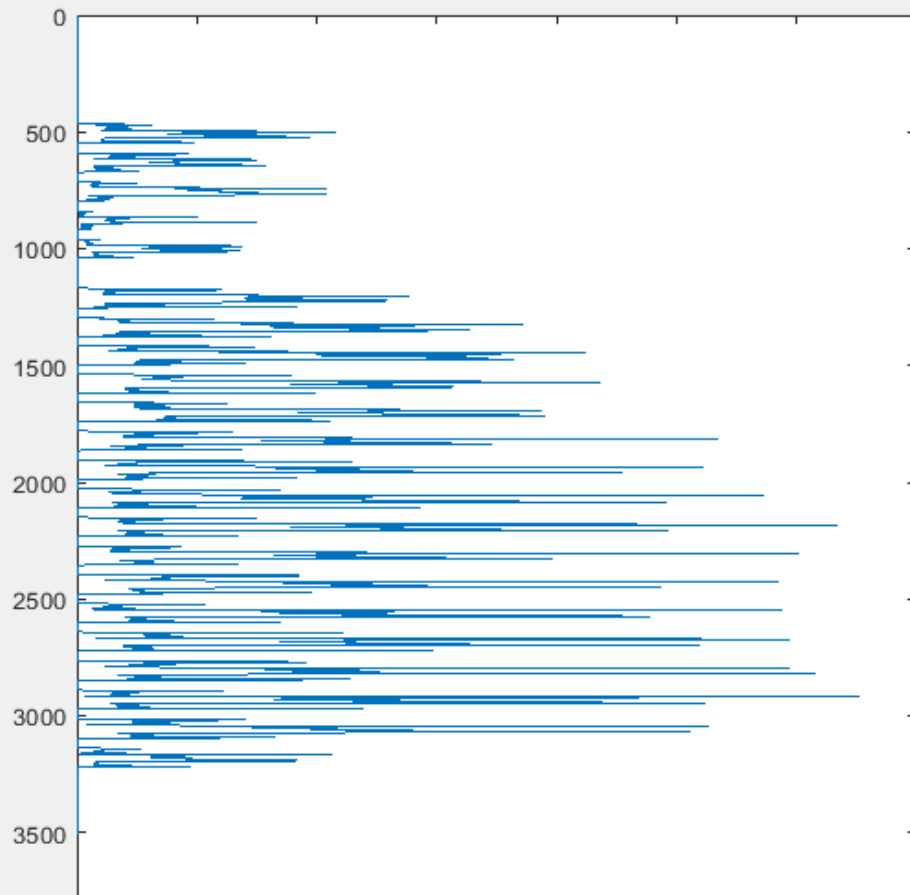
Segmentation

Lines are separated by using horizontal projection profile method.

Character segmentation from segmented lines is done by using vertical projection profile method.

မြို့ပြဆေးကုစုစာပွားကြီးမားခဲ့ပေမယ့်
 နိုင်ငံတော်ဆန္ဒက ဦးစားပြန် ဟု
 ဆရာဝန်များက အခက်အခဲကို ဖော်ပြသော
 သတင်းလွှာ
 (၂၀၁၂ ခုနှစ် မေလ ၁၀ ရက်)

ရန်ကုန်မြို့နယ်အတွင်းရှိ ဘုရား၊ အဆောက်အအုံများကို ချိုးဖျက်ခြင်းများကို ရှိနေသေးသည်။
 ၂၀၁၆ ခုနှစ်တွင် ရန်ကုန်မြို့နယ် အတွင်းရှိ ဘုရား၊ အဆောက်အအုံများကို ချိုးဖျက်ခြင်းများကို ရှိနေသေးသည်။
 ၂၀၁၇ ခုနှစ်တွင် ရန်ကုန်မြို့နယ် အတွင်းရှိ ဘုရား၊ အဆောက်အအုံများကို ချိုးဖျက်ခြင်းများကို ရှိနေသေးသည်။
 ၂၀၁၈ ခုနှစ်တွင် ရန်ကုန်မြို့နယ် အတွင်းရှိ ဘုရား၊ အဆောက်အအုံများကို ချိုးဖျက်ခြင်းများကို ရှိနေသေးသည်။
 ၂၀၁၉ ခုနှစ်တွင် ရန်ကုန်မြို့နယ် အတွင်းရှိ ဘုရား၊ အဆောက်အအုံများကို ချိုးဖျက်ခြင်းများကို ရှိနေသေးသည်။
 ၂၀၂၀ ခုနှစ်တွင် ရန်ကုန်မြို့နယ် အတွင်းရှိ ဘုရား၊ အဆောက်အအုံများကို ချိုးဖျက်ခြင်းများကို ရှိနေသေးသည်။
 ၂၀၂၁ ခုနှစ်တွင် ရန်ကုန်မြို့နယ် အတွင်းရှိ ဘုရား၊ အဆောက်အအုံများကို ချိုးဖျက်ခြင်းများကို ရှိနေသေးသည်။
 ၂၀၂၂ ခုနှစ်တွင် ရန်ကုန်မြို့နယ် အတွင်းရှိ ဘုရား၊ အဆောက်အအုံများကို ချိုးဖျက်ခြင်းများကို ရှိနေသေးသည်။



ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်

နိုင်ငံတော်သမ္မတ ဦးသိန်းစွဲ၏ ထံမှ

အလုပ်သမားနေ့အခမ်းအနားသို့ ပေးပို့သော

သတင်းလွှာ

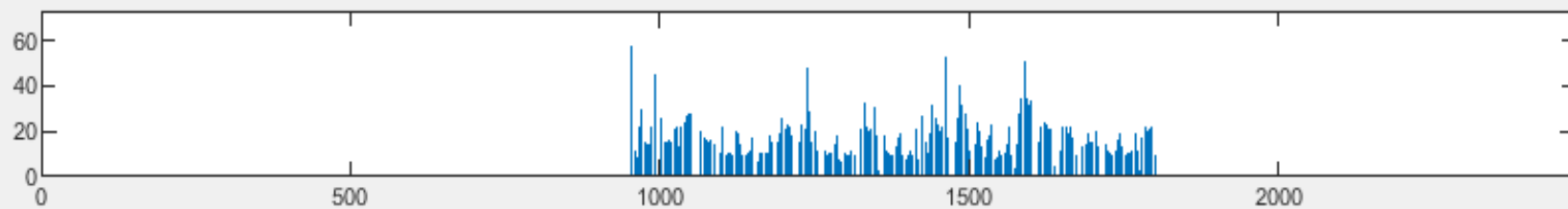
(၂၀၁၉ ခုနှစ်၊ မေလ ၁ ရက်)

ရုတ်တရက်ခလေးစားအပ်သော အလုပ်သမားကြီးများအင်ချား-

ထင်ရှားစွာ ကျောက်လွှာတွင် အလုပ်သမားနေ့အခမ်းအနားကျက် ပြုလုပ်နိုင်ခဲ့အံ့သတိပေး၍
မျောက်ချိုကျောက်လွှာ၌ မြန်မာသမ္မတပြည်အလုပ်သမားများအဖွဲ့အစည်း၏ မြန်မာအလုပ်သမားထု
တစ်ခုလုံးကိုယ်စားပြုရန်အဖြစ်၊ စိတ်ဆုံမျှော်လင့်ခြင်း၊ ပျော်ဝါခြင်း ကောင်းကျိုးပေးလာ
အဖြစ်ဖြစ်နှင့် ပြည်ပကမ္ဘာတစ်ဝှမ်း လက်ကိုင်အမှုဆောင်သမားများ၊ လူမှုစီးပွားဝန်ထမ်း၊ မြို့ခြံ
တိုးတက်ကြံ့ခိုင်ရေးနှင့် ဖွံ့ဖြိုးရေးကော်မရှင်များအဖွဲ့၊ ဦးစွာနှုတ်ခွန်းဆက်သအပ်ပါသည်။

ကမ္ဘာတစ်ဝှမ်းရှိ အလုပ်သမားများကို အလေးထားကူညီပြုသည့်အနေဖြင့် ပေါ်ပေါက်
ခဲ့သော အလုပ်သမားနေ့သည် မေလ ၁ ရက်နေ့အားဖြင့် ရည်ရွယ်သည့် သမိုင်းအကြောင်းနှင့်
အကျိုးကျေးဇူးတစ်ခုရရှိခဲ့ပြီဖြစ်ပါသည်။ စက်မှုထုတ်လုပ်ရေးအဆက်ဆက်တွင် အလုပ်သမား
များအနေနှင့် ဂုဏ်ရည်ထက်သန်သည့် အခွင့်အရေးများ၊ လိုလားချက်များကို ပစ်ခန်းမခံ
မျိုးစုံကြားမှ တောင်းဆိုရလျှင်ဖြစ်ကြပါသည်။ လက်ရှိ ကာလတွင် မြန်မာနိုင်ငံအပါအဝင်
ကမ္ဘာတစ်ဝန်းမှ အလုပ်သမားများ၏ ရသင့်ရတိုက်သည့် အခွင့်အရေးများသည်
လာစဉ်လာက ပိုမိုတိုးတက်ရန်အတွက် တောမြင်နိုင်ပါသည်။ သူတို့သို့ ကိုးဟပ်မှုများ
ပြုစုပေးလာရန် စိတ်ထားပေးရန်နှင့် ဦးစွာကူညီခြင်းသည် အလုပ်သမားများကို
လေးစားသတိထားပြုအပ်ပါသည်။ အလုပ်သမားများ ရသင့်ရတိုက်သည့် အခွင့်အရေး
များကို ဥပဒေနှင့်အညီ ရရှိခံစားနိုင်ရန် နိုင်ငံတော်အစိုးရအနေဖြင့် အလေးထား
ဆောင်ရွက်ပေးလျက်ရှိပါသည်။

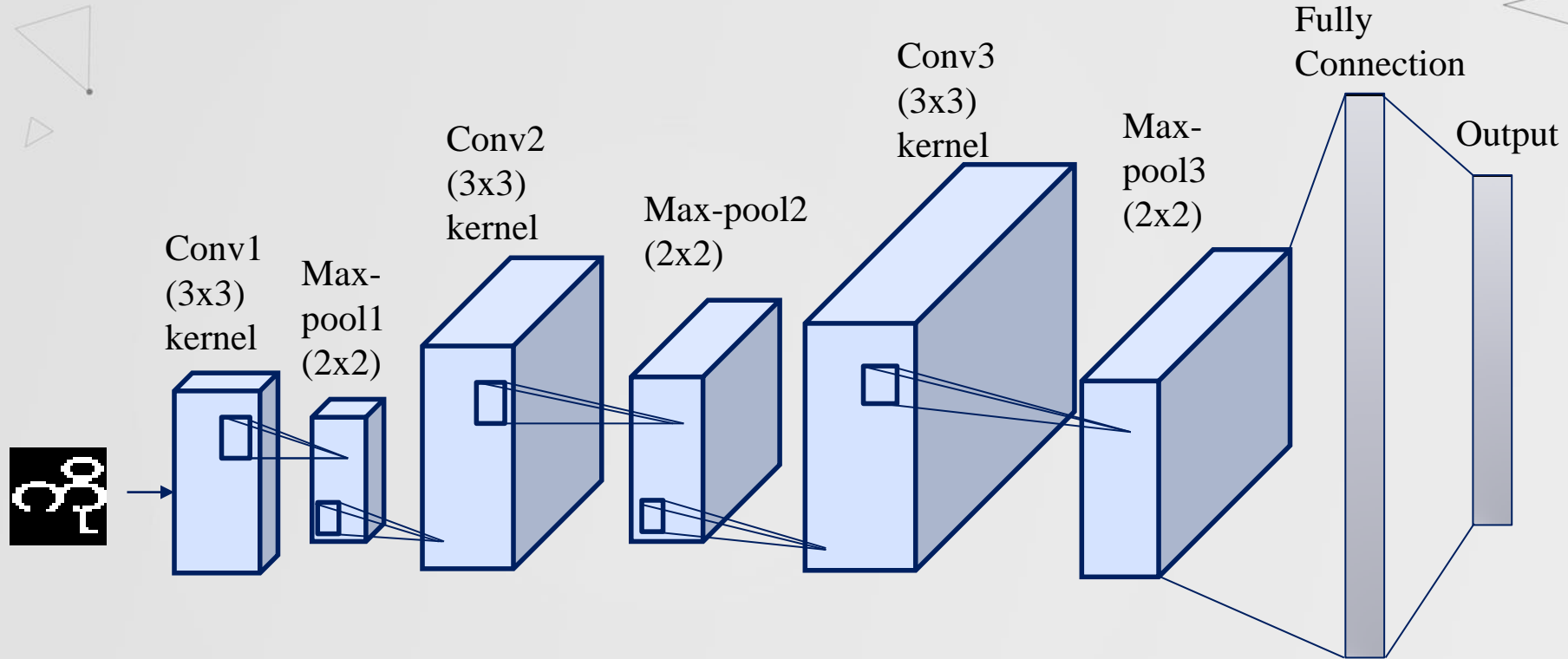
ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်


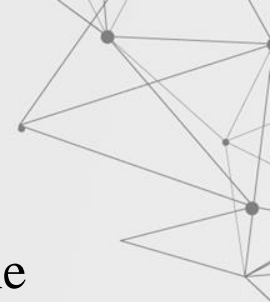


ပြည်	ထောင်	စု	သမ္မတ	မြန်မာ	နိုင်ငံ	တော်
ပြည်	ထောင်	စု	သမ္မတ	မြန်မာ	နိုင်ငံ	တော်
ပြည်	ထောင်	စု	သမ္မတ	မြန်မာ	နိုင်ငံ	တော်

Recognition

It is performed by using the pre-trained Convolutional Neural Network (CNN).



- 
- 
- This architecture is used for both training and testing the Myanmar Character recognition.
 - Input Layer - (32x32x1) image
 - Three convolutional layers with $n1=64$, $n2=128$, and $n3=256$ filters respectively
 - Batch normalization layer between Conv layer and ReLU
 - Max Pooling after each ReLU layer
 - FC (i.e. fully-connected) layer with a softmax activation function in the output layer (583classes)

CNN Training

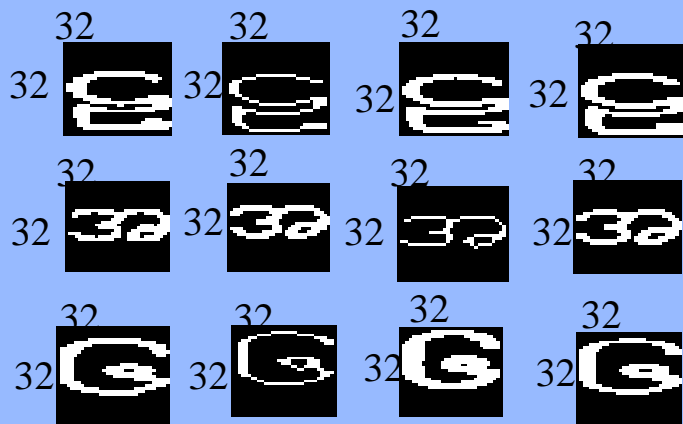
We trained the CNN for total 583 classes :

- Consonants = 33 classes ,
- Digits = 10 classes, and
- Compound words like – “ကံ”, “ကျ”,
“ကာ”, “ကို” = 540 classes
- Total images = 43725

For each character, we use:

- Input Size = 32x32
- Training data = 60 images
- Testing data = 15 images
- Font size = 12, 14, 16, 20, 24

Compound words



Consonants and digit

Post-processing

The recognized characters are converted into Unicode and saved in a text file.

There is one of the complexities of implementing Myanmar script in Unicode:

နေ	န	ေ	
	1014	1031	the sun

ပေါ	ပ	ေ	ါ	
	1015	1031	102B	plentiful

ကျောင်း	က	ျ	ေ	ွ	င	း	
	1000	103B	1031	102C	1004	103A	1038 school

Challenges in OCR

- Font styles and sizes, mixed scripts.
- If the original source image itself is blurring and has poor resolution, the recognition can't be good.
- Accuracy of OCR heavily depends upon segmentation phase.
- The recognition also fails for the characters which have close resemblance.





Conclusion

- **OCR**
- **Usage of OCR**
- **Challenges of OCR**

Futther Extensions

- **Hand written recognition**
 - **Pali recognition**
 - **Character correction**
- 

Thank you!

