Optical Character Recognition for Myanmar Language

Khin Khant Khant Hlaing

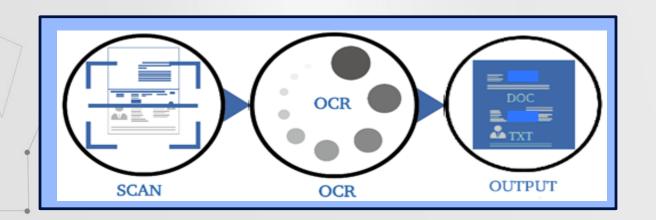
Outline

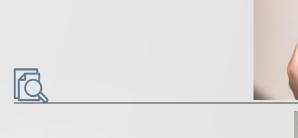
- > What is OCR?
- > Myanmar Script
- > Implementation of the Myanmar OCR



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.





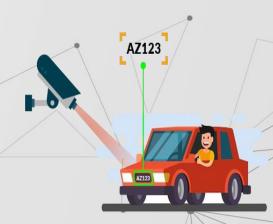
Some usage of the OCR system

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

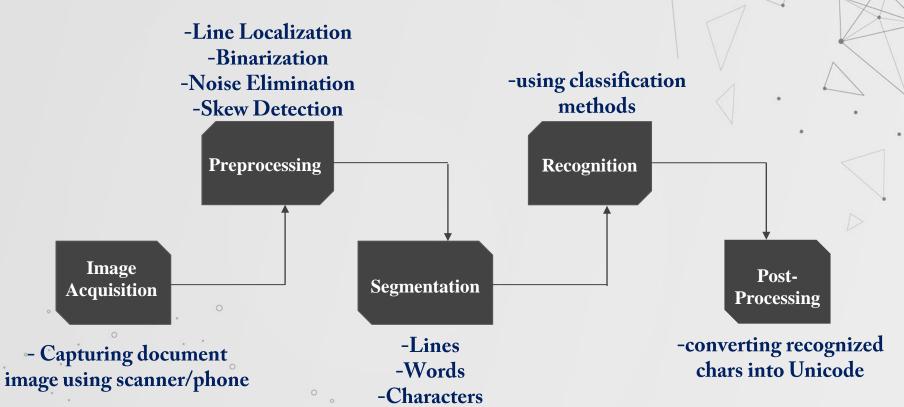








Steps in OCR



Myanmar Script

33 consonants:	ကခဂဃငစဆဲ စစ္နည္ရင္ခ်ဥ္ၿပကတဲ့ တဒဓနပဖဗဘမယရလဝသဟဋ္ဌအ
Numbers:	၀၁၂၃၄၅၆၇၈၉
Vowels:	-> -J = - = - = - = - = - = - = - = -
Medial:	16-1
Independent vowels:	က္က ဤဥဦ ဧဩဪ
Symbols and punctuation:	၌ ၍ ၎င်း၏ ၊ ။

Figure. A set of Myanmar Alphabet

- 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.

Implementation of the Myanmar OCR

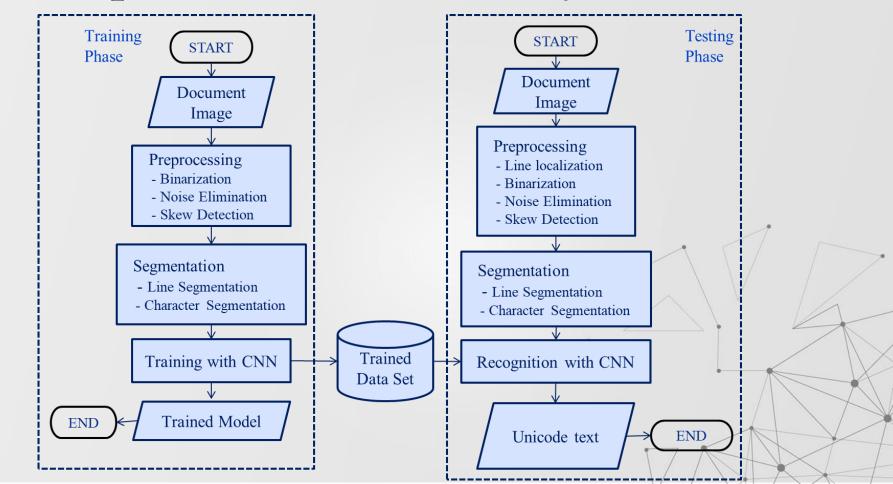


Image Acquisition

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

ချစ်ကြည်လေးစားအပ်ပါသော အလုပ်သမားကြီးများခင်ဗျား–

ယနေ့ကျရောက်သည့် အလုပ်သမားနေ့အခါသမယမှာ ပြည်ပနိုင်ငံအသီးသီး၌ ရောက်ရှိနေကြသည့် မြန်မာရွှေ့ပြောင်းအလုပ်သမားများအပါအဝင် မြန်မာအလုပ်သမားထု တစ်ရပ်လုံး ကိုယ်၏ကျန်းမာခြင်း၊ စိတ်၏ချမ်းသာခြင်း ပြည့်ဝပြီး ကောင်းကျိုးမင်္ဂလာ အဖြာဖြာနှင့် ပြည့်စုံကြပါစေ၊ လုပ်ငန်းခွင်အေးချမ်းသာယာပြီး လူမှုစီးပွားဘဝများ ဖွံဖြိုး တိုးတက်ကြပါစေလို့ ဆုတောင်းမေတ္တာ့ပို့သရင်း ဦးစွာနှတ်ခွန်းဆက်သအပ်ပါသည်။

ကမ္ဘာတစ်ဝန်းရှိ အလုပ်သမားများကို အလေးထားဂုဏ်ပြုသည့်အနေဖြင့် ပေါ်ပေါက် ခဲ့သော အလုပ်သမားနေ့သည် ခေတ်အဆက်ဆက် ရှည်လျားသည့် သမိုင်းကြောင်းနှင့် အတူ ကျော်ဖြတ်ခဲ့ရခြင်းဖြစ်ပါသည်။ စက်မှုတော်လှန်ရေးအဆက်ဆက်တွင် အလုပ်သမား များအနေနှင့် ရသင့်ရထိုက်သည့် အခွင့်အရေးများ၊ လိုလားချက်များကို အခက်အခဲ မျိုးစုံကြားမှ တောင်းဆိုရယူနိုင်ခဲ့ကြပါသည်။ လက်ရှိ ကာလတွင် မြန်မာနိုင်ငံအပါအဝင် ကမ္ဘာတစ်ဝန်းမှ အလုပ်သမားများ၏ ရသင့်ရထိုက်သည့် အခွင့်အရေးများသည် ယခင်ကထက် ပိုမိုတိုးတက်ရရှိနေကြောင်း တွေ့မြင်နိုင်ပါသည်။ ယခုကဲ့သို့ တိုးတက်မှုများ ဖြစ်ပေါ်လာရန် စိတ်ထားကောင်းနှင့် ဦးဆောင်ဦးရွက်ပြုခဲ့သည့် အလုပ်သမားများကို လေးစားအသိအမှတ်ပြုအပ်ပါသည်။ အလုပ်သမားများ ရသင့်ရထိုက်သည့် အခွင့်အရေး များကို ဥပဒေနှင့်အညီ ရရှိခံစားနိုင်ရန် နိုင်ငံတော်အစိုးရအနေဖြင့် အလေးထား ဆောင်ရွက်ပေးလျက်ရှိပါသည်။

Preprocessing

Removing logo by using the saturation in place of intensity.

Otsu's algorithm is used for binarization.

Median filter is used for noise elimination.

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

ပြည်သောင်စုသမ္မတဖြစ်မာနိုင်ငံတော် နိုင်ငံစောဒိသမ္မတ ဦးဝင်းမြင့် သံမှ အကုပ်သမား နေ့အခမ်းဆနေသသို့ ပေးပို့ယော သဝတ်လွှာ (၂၀၁၉ ခုချစ်၊ လေ ၁ ရက်)

ချစ်ကြည်လေစားအပ်ပါသော အလုပ်သမားကြီးများစစ်ဗျား-

းခန္းျမရာတိသည့် စာသပ်သမားနေ့အခါသမာမှာ ပြည်းခိုစိုင်းမေးပီးလိုင်း ရောက်ရိုင်နကြသည့် မြန်မာရွေ့မြောင်းအလုပ်သမားများစာပါတဝင် မြန်မာအလုပ်သမားလု တစ်ရပိုလုံး ကိုယ်၏ထုန်းမာမြင်း၊ ခိုက်၏ခုခိုးလာခြင်း မြည့်ပြုပြီး လောင်းကိုးဖပ်သာ အမြင်များနှင့် ပြည့်စုံကြုပ်မေး လုပ်ပန်းမှုင်သေးရစ်သေားသင်းစုံ လုမ္ပတီပွားကရေးသည့်မြိုး ကိုးသည်ပြည်းမေးကို သူတောင်းမွေတွဲသူလိုသခင်း ဦးစွာနေတိခွန်းထက်သံအစိုပါသည်။

ကန္တာတစ်ခန်းရှိ အသုပ်သမားရာတပို အလေးကာကျကီးပြာသည့်အနေဖြင့် အခြင်းကို ခဲ့သော အလုပ်သမားနေသည် အသေးအဆက်ဆက် ရှည်လွာသားလို အခြင်းကြောင်းနှင့် အဘု ကျော်ဖြတ်ပုံရခြင်းဖြစ်ပါသည်။ မက်ခဲ့တော်လှန်းဆုံးဆောက်ဆက်တွင် အသုပ်သမား များမေနနှင့် သေခိုရတိုက်သည့် အချင်အရေးမှား၊ လိုလားချက်များကို အကော်အခဲ မျိုးစီပြောမှ မဟာင်းဆုံးပုသနိုင်ခဲ့ကြပါသည်။ သတ်ရှိ ကာလတွင်းမြန်မာနိုင်ငံအပါအကြီး တန္တာတစ်ပန်းမှ မလျှင်သမားမွား၏ ခုသလိုစာဆိုက်သည့် အချင်အခွဲရေးများသည် ယခင်တယက် ပိုစ်တိုတာကိုးရှိနေကြောင်း တွေ့မြင်နိုင်ပါသည်။ တခုကိုသို့ တိုကောင်မှုခုအ ဖြစ်ပေါ်သာရန် စိတ်ထားသကင်းနှင့် ဦးသောပင်ဦးရွယ်ပြန်သည့် အလုပ်သမားများကို လေးသေးလေသမှတ်ပြုဆင်ပါသည်။ အလုပ်သမားများ ရသင်းလိုက်သည့် အချင်အစေန ပျားလို ဥပဒေနှင့်အညီ ရရှိခံစားနှင့်တြန် နိုင်ငံတော်အခိုးရာအနေဖြင့် အထောယား ဆောင်ချက်လေးကွက်ရှိခါသည်။ ခြဲသို့အောင်ခံခဲ့သည့်က ခြဲရိမာနိုင်စေလို နိုင်ငံတော်သမှုက ဦးလုံးမြဲခို ကိုမှ သယုပ်သမားရေ့အခမ်းအအလော့ လေးမို့ကော သမာဇ်တူဝ (၂၀၁၈ သည်။ မခုလ (၁) ရက်()

ချ**စ်**ကြည်ထေးစေးအင်းမသော အလည်းသမားကြီးများဆင်ဗျားမ

သားနှာ များရာကိုသည့် အလုပ်လေးမှာ နားခါသားမှာ ေပြည်သွင့် စီးမွာမီးကို၌ ရောက်ရှိမှသြောသည့် ဖြန့်မားရှာမြောင်းအလုပ်လာစားများအပါတလို ဖြန်းကလုပ်လာလည်း ကင်ရပိုင်မှာ ကိုယ်မှာရသော ဖြစ်။ ဒီလိုယ်ခု မီးလာခြင်း ပြည့်ခြင်း ကောင်းလင်းပင်လာ အချိန်ကုန် ပြည့်ခဲ့ကြပါသေ လုပ်ပန်းမွှင်သားရခန်းသာလာရှိသေ့မှစ်ရွားလာစနား ရှိရှိခဲ့ ထိုလာကျသည်ကောင့် ကျလောင်းလတ္တာရှိသည်။ ဦးမွာနေပါမှုနီးသားပါသားမိပါသည်။

ကန္တာလော်လနိုင်ရှိ အလုပ်သလားမွားကို အတေလလာလူတီပြားသည်။ လေဖြင့် လပြတ်ကို မိုင်သာ အလုပ်သလောလည်း လေသိ ယောက်သားကို ရှည်ကျာလာတဲ့ လမိုင်းကြောင်းနှင့် အဟု လေခွင်ပြာင်ခဲ့ရခြဲခဲ့ရသော လည်းသောကောက် ရှည်ကျာလာတီတွင် အလုပ်သလေ များအရေနှင့် စာယင်ရသိုက်သည်။ လည်းသောက်မှ နိုလာလချက်များကို အာလာတစ် မိုးစုံပြုပြည်မှ မေသင်းဆိုရသနိုင်ရပြေပါလည်း။ လက်စို ကလာတစဉ် မြန်မာနိုင်ငံအပါအသင် လည်းသည်။ မိန်းမှ မေသင်းဆိုရသနိုင်ရပြေပါလည်း။ လက်စို ကလာတစဉ် မြန်မာနိုင်ငံအပါအသင် လည်းသည်။ မိုးမိန်းမှ မေသင်းမြန်းမြောက်မြောင်းမှ တွေမြင်ခိုင်ပါသည်။ လနာကိုလို တိုလောင်မေလျှင် မြန်းပြေသည်။ မိုးမိုးမေလေးများကောင်းမှုနှင့် ဦးစေတင်ခိုးရှာသိပြုခဲ့သည့် အလုပ်သမားရေးကို လေလေးသည်။ များပြုပြေသင်းကြသည်။ မာလုပ်သမားရေး ရသည်စာသို့ လေချင်ပြင် အလေလသင်း ရောက်ချင်းမှာလျှတ်ရှိ ပါသည်။



ပြည်လေးခဲ့ရသည္ ကမြန်မာနိုင်ငံလေး နိုင်ငံးတစ်သမ္မက ဦ(၁)() ရှိမဲ့ ယုံမှ အလုပ်သမားရေးအခင်းအများကို မေးရှိသေး သခတ်သွှင ()မာမှ မှနှစ် ရေးပ (၁) ရက်)

ရှင်ကြည်းကေရာအခင်ပါတယာ အလုပ်သမား ကြားများကိုများ

గా కెళ్ళువుకా లే ఇంత్ర్లో అక్కాని అండు కార్యా ఇదా ముందులో ప్రైవేంధి వరణుచేటుకున్న అయిన అంట్రామై అక్కి అంట్లు అధ్యా వియాది ఇదా అయిని అర్వారికి అయిని అన్ని అయిని అన్ని అయిని అన్ని అయిని అన్ని అ అంట్లు అంట్లు అంట్లు అంట్లు అంట్లు కూడి అంట్లు అ అంట్లు అంట్లు అంట్లు అంటు అంట్లు అంటు అంట్లు అంట్లు

မန္တား လိုင်ရိုးရှိ အကုန်လာမှာႏွစ်ကို ဘလေးလာရာလီပြင်လျှင်စာနေပြီး ပေါက်လို စို့သော အလုဒ်သားသာရနည်းကို ရေးခံသနာကို အကို ရှည်လူ မတည့် သန်ပိုကြောင်းမှုရှိ ကျေး သည်ခြင်းရြေးကြောင်းမျှင်ချီသည်။ မလုပ်မှာရေးများ၊ ကိုလာရာရာစီများကို အသားမေး ပြောရိုကျွန်းမှ ကောင်းပေးလှုန်ခန့် ပြုပ်သည်။ လာလို့ရလုံးပေးညို အနှင့်ရှိမပြောမိုပိုင်သေးပြန်သင့် တည်သားမှာနှင့် အလင်းသမားများ၏ ရသန်ရလုံးကိုလည်း ငေးခံထိုလို အိုးရှိခဲ့ရေးနေရာမှုနှစ်ပြီးမှာနိုင် သည်လာလင် ပိုနီးဆိုင်သည် မျှင်မေးကြောင်း တွေ ခြင်းနီးခဲ့သည်။ ငေးခံထဲလို ကို အားကိုမှောင် မြန်မေါင်လာရန် စိတ်ထားတောက်သည် ဦး ဆောင်ဦးချက်ပြန်လည်း ကလုံကလာခဲ့ချားကို ကလေးကောက်လင်းပြုအပြေးလည်း အင်းသည်။ သင်းမှာများနှင့် လောင်လည်းလိုင်သည့် အနှင့်အနေနေ များလို့ ဥပဒေ မေရိသနည်း ရခြန်သားနှင့်စုံရန် စိတ်စံသစ်အစီးမေးမှာပြေပြီး အလေးသည်း ယောင်ရုံး စိုးလေးလျက်ပြေးသည်။



ပြည်လောဂ်စုသျှဘမြန်မာနိုင်ငံတော် နိုင်ငံအက်သမှုတ ဦးဝင်ပြင် ထံမှ အလုပ်သမားနေ့အခမ်းအနားတို့ ပေးပိုသေး သဝတ်လွှာ (၂၀၁) မှနှစ် မေလ ၁ ရက်)

ရပ်ကြည်းကေအားအပ်ပါသေး အလုဒ်သီးမှာ ကြီးများစင်ဈားမ

ကနေ့ကျနောက်သည့် အလုပ်ယမားအနှစာင်သမားသေ ပြည်ဝန်ခံစံသည်းသီး၌ ခရာလ်ပို့ရေးသို့သည့် မြန်မာရေ့မြောင်းအလုပ်သမားများအပါအဝင် ဖြန်မာကလုပ်သမားကို တစ်ရုပ်လုံး ကိုယ်၏မျှနိုင်မာဖြစ်၊ ဒီကို၏ချမ်းကာဖြင်း ပြည့်ဝပြီး တောင်းကျွဲးခင်သင သမြေမှနှင့် ပြည့်ခဲ့ကြပါစစ လုပ်ငန်းခွင်အေးချမ်းသားသေးနဲ့ လူမှုခီးပွားကလနာ ရှိပြီး ပရိုးမာက်ကြပ် စလေ့ နာတောင်းရေကွာျှဉ်သရကီး ဦးခွာနှက်ခွန်းဆက်သနာပေါ်သည်။

ာနှားပစ်ခန်းရှိ အကုဂ်လာအနာကို အလေးလားခဲ့တို့ ပြည်ညိုစာနေပြင့် ပေါက်လို နိမ်သာ အလုပ်သပ်လေနလည်း မောက်အနာကို အလုံးသည် သနိုင်ကြောင်းနှင့် တွေ သည်ခြင်းခြင်းပြည်ခြင်းပြည်။ မောင်အနာကို မောင်လှန်းမှာလေးပို့သတည်း သနိုင်ကြောင်းနှင့် တွေ သည်ချင်းကြောင်းနှင့် အမှုင်းမေးများ သည်ချင်းကြောင်းမှန် ပြေပါသည်။ လက်ရှိ ပေသာတွင် မြန်မာနိုင်ငံအပါအပင် ကမ္ဘာ့တစ်ပန်းမှ အလုပ်သမားများ၏ ရသင့်ရသိုက်သည့် အနှင့်အရေးများသည် သင်လာလက် ပိန်တီးသည်းနှိုမေကြောင်း တွေ့ခြင်နိုင်ငံတွေ၊ လောက်လုံးကို တက်မှုမှာ ဖြစ်ပေါ်လာရန် စိတ်သေးကောင်းနှင့် ဦးသောင်ဦးရွာက်ပြည်သည့် အလုပ်သမားများကို လာစာကောင်ကြောင်ပြုအပြေသည့်။ အလုပ်သမားများ ကောင်လုံးလိုက်သည့် အနှင့်အနေးမှ ပညာပြုံ ဥပဒေမှန်အသည့် ရရှိခံသင်္ခေဝရန် စိတ်လောင်အနှင့် မေးသပ်ရှင်လည်း လောင်လုံးလည့် အနှင့်အနေး ပညာပြုံ ဥပဒေမှန်အသည့် ရရှိခံသင်္ခေဝရန် စိတ်လောင်အစိုးအေခနေဖြင့် အလောလာသင်းသင့်မှု ပို့ပြောင်ပြုအပြေသည့်။ ရရှိခံသင်္ခေဝရန် စိတ်လောင်အစိုးအေခနေဖြင့် အလောလာသင်းမေးပြုံးကိုသည့်။ ရရှိခံသင်္ကေဝရန် စိတ်လောင်အစိုးအေခနေဖြင့် အလောလာသင်းမေးပေးမှုကိုရှိပေးကျက်ခြင်းသည်။



^{ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်} နိုင်ငံတော်သမ္မတ ရွိ^နဝင်းမြင့် ထံမှ အလုပ်သမားနေ့အခမ်းအနားသို့ ပေးပိုသော _{మరియ}్తు

(၂၀၁၉ ခုနှစ်၊ မေလ ၁ ရက်)

ဆောင်ရွက်ပေးလျက်ရှိပါသည်။

ချစ်ကြည်လေးစားအပ်ပါသော အလုပ်သဝေးကြီးများခင်ဗျား– ယနေ့ကျရောက်သည့် ဆလုပ်သမားရန္ အခါသမယမှာ ပြည်ပနိုင်ငံအသီးသီး၌ ရောက်ရှိနေကြသည့် မြန်မာရွှေ့ပြောင်အလုပ်သမားများအပါအဝင် မြန်မာအလုပ်သမားထု တစ်ရပ်လုံး ကိုယ်၏ကျန်းမာခြင်း၊ စိတ်၏ချမ်းသာခြင်း ပြည့်ဝငြီး တောင်းကျိုးမင်္ဂလာ အဖြာဖြာနှင့် ပြည့်နုံကြပါစေ၊ လုပ်ငန်းခွင်အေးချမ်းသာယာပြီး လူမှုစီးပွားဘဝမှုအ နွံဖြုံး တိုးတက်ကြပါခဲ့လေ့ ဆုတောင်းမွေတွာပို့သရင်း ဦးမွာနှတ်ခွန်းဆက်သအစ်ပါသည်။ ကမ္လာတစ်ဝန်းရှိ အလုပ်သမားများကို အလေးထားဂုတ်ပြသည့်အခနဖြင့် ပေါ်ဝေါက် ခဲ့သော အလုပ်သမှသနေ့သည် ခေတ်အဆက်ဆက် ရှည်လျားသည့် သမိုင်းကြောင်းနှင့် အတူ ကျော်ဖြတ်ခဲ့ရဖြင်းဖြစ်ပါသည်။ စက်မှုတော်လှန်ရေးအဓာက်ဆက်တွင် အလုပ်သမား များအနေနှင့် ရသင့်ရထိုက်သည့် အခွင့်အရေးများ၊ လိုလားချက်များကို အခက်အခဲ ရိုးစုံကြားမှ တောင်းဆိုရယူနိုင်ခဲ့ကြပါသည်။ လက်ရှိ ကာလတွင် မြန်မာနိုင်ငံအပါအဝင် ကမ္ဘာတစ်ဝန်းမှ အလုပ်သမားများ၏ ရသင့်ရထိုက်သည့် အခွင့်အရေးများသည် ယခင်ကထက် ဂိုနီတိုးတက်ရရှိနေတြောင်း တွေ့မြင်နိုင်ပါသည်။ ယခုကဲ့လို တိုးတက်မှုများ ဖြစ်ပေါ်လာရန် စိတ်ထားကောင်းနှင့် ဦးဆောင်ဦးရွက်ပြခဲ့သည့် အလုပ်သမားများကို လေးစားအသိအမှတ်ပြုဆင်ပါသည်။ ဆလုပ်သမားများ ရသင့်ရတိုက်သည့် အခွင့်အရေး များကို ဥပဒေနှင့်အညီ ရရှိခံစားနိုင်ရန် နိုင်ငံတော်အစိုးရအနေဖြင့် အလေးထား

ြည့်မှ ပျပိရလမှုလဖြန့်**တရိုင်ငံတော်** နိုင်ငံတော်သမ္မတ ဦးဝင်းခြင့် ကုန အတူစီသမားနေ့နေရန်ကနေနည်း အချိန်သည non-figge ါ့လ**ှဲ ရှေစီ စေ**သ ၁ ကြီး

ခုနီကြည်လေးကောက်ပါတော့ အတွင်သဟာကြီးများစင်များ

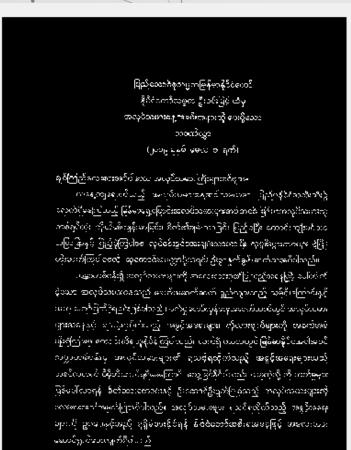
ကောက္ ကေက်ာက္၍ အလုပ်သတာရေ အခ်ကလေမာင္းကြက္ခ်ပိုင္းသိုက်ခဲ့ းရက်ရှိရေကြသည့် မြို့သည့် ပြောင်းသည်လေးကျင်လည်းကို မြန်မာတွင် လျှင်လျှင် အက်လုံးသုံး ကိုလ်ကိုကျွန်းမှ ပြုပါ။ နဲ့ ပိန်းမှုပို့လကျွန်း၊ ပြုညှိရခြီး နေတစ်းကျိုပင်လာ အဖြဲးရိုးရနှံ့ မြှုန့်နှံမြေးကေး လုပ်နေ့နေရသောအနှင့်သာလာသီး အမှုနီးမျှသတများ ခွဲခြင့ ကိုလာတိုင်ကြောမှာတို့ ကျလောင်းပေးသွားမို့သာကိုလည်း ပြန္လည္မကိုရနီးဆက်လာလိုပါသည်။

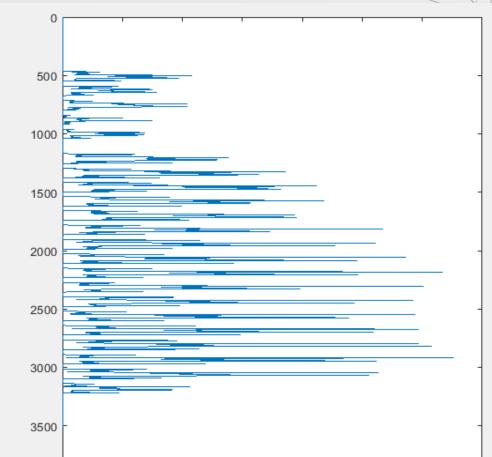
ကန္တာကစိုင္ရနီရနီ အလုပ်သည္။ ဥပဂၢို အသည္။ ေျပစိုင္ခ်န္း ရဲ႔လုပ္ျပည္ခဲ့ အခြဲေပါ စုံသော အားသိုးသမား နှားကို လက်သောကိုဆက် မျှည်ကျားသည့် သင့်ပိုအပြင့်စပ်နှင့် ေကျမတည်ရှင်ကိုရန်ဦးမြင်ပါသည်။ စက်မှုလေးပြည်းနောက်သက်တွင် အလုပ်သတင များအနေနှင့် ရသင့်ရလိုက်သည်။ အတွင် ကေးသူသူ၊ လိုက်ရေးကိုများများ၊ အချက်မည် $\frac{d^2}{dt}\sqrt{2\gamma}$, λ_{ϕ} and the $\frac{d}{dt}$ -profit of $\frac{d}{dt}$ is each constant of the second state ာတ္သား ဝပိုဝရိုရေ။ အာတ္ပမိုင္အာမေရးများ၏။ ရသင္ခ်ရဆိုက်သည့္တဲ့ အာရွင္ခ်မစေရးများသည်။ we have only lightly and object, fixed a manifold for $(0,\infty)$ by a fixed and $(0,\infty)$ ကြိန်လျှင်း လူ၍ ကိုတ်လာလေး စီးနှင့်။ ဦးဆောင်ဦးရွက်မြို့နာသည့် အတွင်အထားများကို $we wanted and \{[a-b], a_i\}_{i=1}^n = (ab) consequence \{[a,b], p^i\} be (a) any boxes.$ ရှားမရှိ ဥပဒေနှင့်အ<u>င်</u>၌ ဝရှိစတ^{ှန့်}မိရန် သိနိန်တက်ကန်းရအနှန့်နှင့် အလေးအား $asset \delta_{ij} absorber ab fill a g \hat{\alpha}_{ij}$

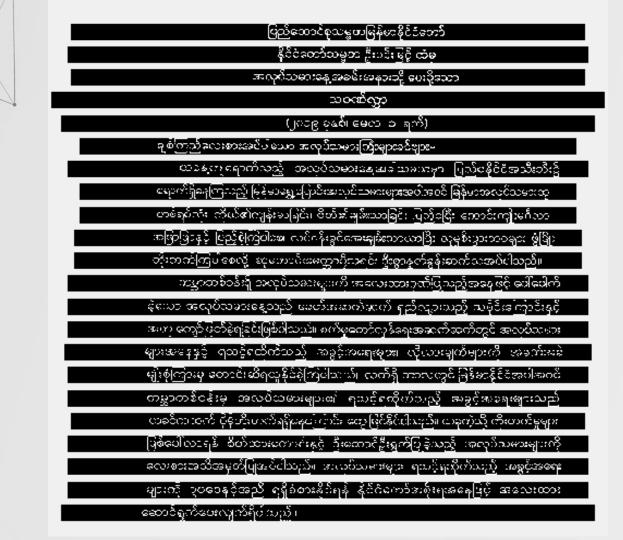
Segmentation

Lines are separated by using horizontal projection profile method.

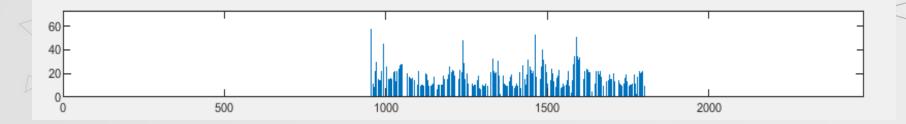
Character segmentation from segmented lines is done by using vertical projection profile method.

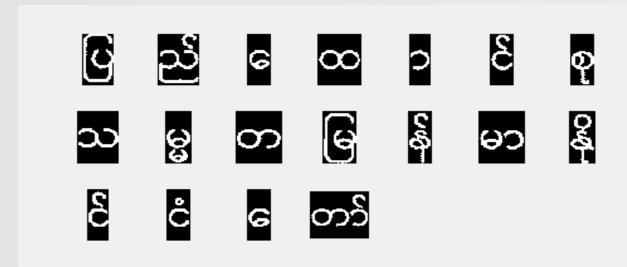






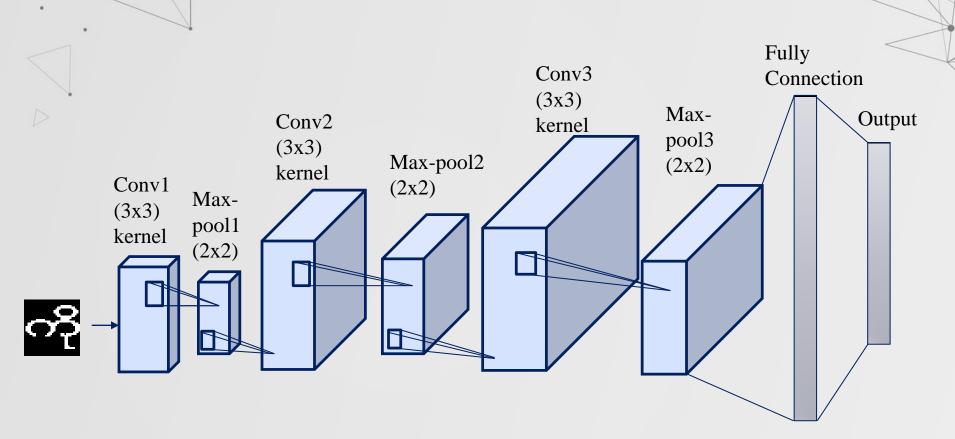
ပြည်ထောင်စုသမ္မတမြန်မာနိုင်ငံတော်





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 "က်", "ကျ", ိ"ကာ", "ကုံ" = 540classes
- \triangleright Total images = 43725

For each character, we use:

- ightharpoonup Input Size = 32x32
- \triangleright Training data = 60 images
- \triangleright Testing data = 15 images
- \triangleright 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	ه 1014	6 1031		the	sun			
ပေါ	o 1015	6 1031	ົງ 102B	ple	entiful			
ကျောင်း	თ 1000	↓ 103B	6 1031	o 102C	c 1004	c 103A	: 1038	school

Challenges in OCR

- **3**
 - > 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

