# Watermarking Text Documents With Watermarked Fonts

C. He, D. Wu, X. Zhang and Hanzhou Wu

Shanghai University

ACM IH&MMSec'24, Vigo, Spain

# Outline

1 Motivation

2 Proposed Method

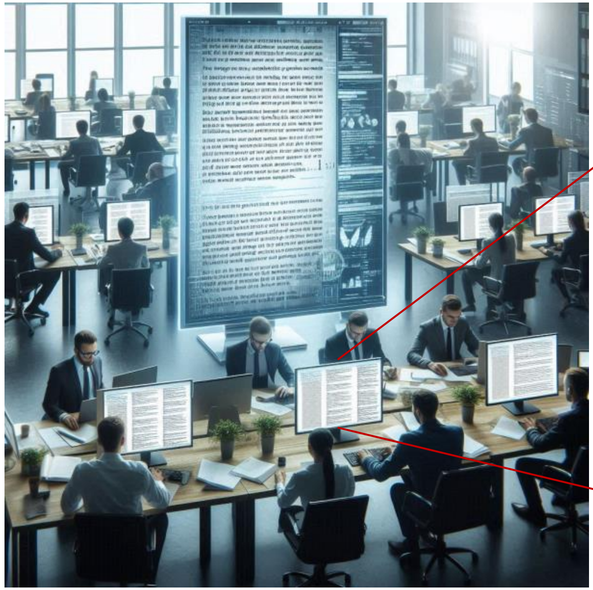
3 Experimental Results and Analysis

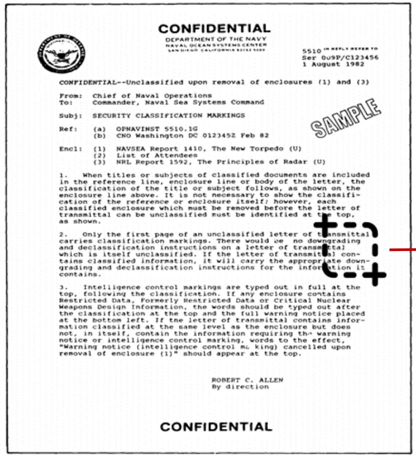
4 Conclusion and Discussion

# 1 Motivation

## Problem

- **How to prevent sensitive text contents from being leaked by screenshots by *digital watermarking*?**
- Comes from *Tencent Inc. (project cooperation)*





CONFIDENTIAL
DEPARTMENT OF THE NAVY
NAVAL OCEAN SYSTEMS CENTER
SAN DIEGO CALIFORNIA 92152-5000

5510 IN REPLY REFER TO
Ser 009P/C123456
1 August 1982

CONFIDENTIAL--Unclassified upon removal of enclosures (1) and (3)

From: Chief of Naval Operations
To: Commander, Naval Sea Systems Command

Subj: SECURITY CLASSIFICATION MARKINGS

Ref: (a) OPNAVINST 5510.1G
(b) CNO Washington DC 012345Z Feb 82

Encl: (1) NAVSEA Report 1410, The New Torpedo (U)
(2) List of Attendees
(3) NRL Report 1592, The Principles of Radar (U)

SAMPLE

1. When titles or subjects of classified documents are included in the reference line, enclosure line or body of the letter, the classification of the title or subject follows, as shown on the enclosure line above. It is not necessary to show the classification of the reference or enclosure itself; however, each classified enclosure which must be removed before the letter of transmittal can be unclassified must be identified at the top, as shown.

2. Only the first page of an unclassified letter of transmittal carries classification markings. There would be no downgrading and declassification instructions on a letter of transmittal which is itself unclassified. If the letter of transmittal contains classified information, it will carry the appropriate downgrading and declassification instructions for the information it contains.

3. Intelligence control markings are typed out in full at the top, following the classification. If any enclosure contains Restricted Data, Formerly Restricted Data or Critical Nuclear Weapons Design Information, the words should be typed out after the classification at the top and the full warning notice placed at the bottom left. If the letter of transmittal contains information classified at the same level as the enclosure but does not, in itself, contain the information requiring the warning notice or intelligence control marking, words to the effect, "Warning notice (intelligence control marking) cancelled upon removal of enclosure (1)" should appear at the top.

ROBERT C. ALLEN
By direction

CONFIDENTIAL

Sensitive content

Screenshot

Leakage

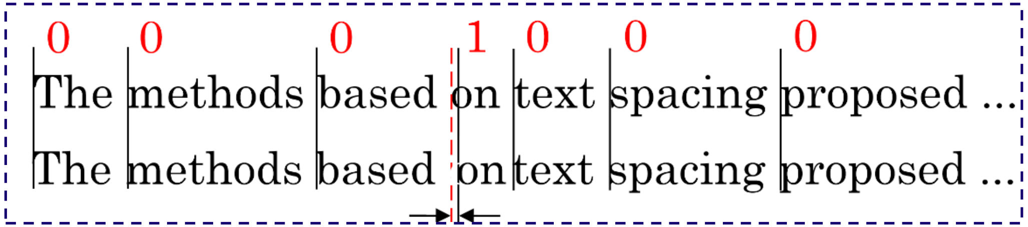
# 1 Motivation

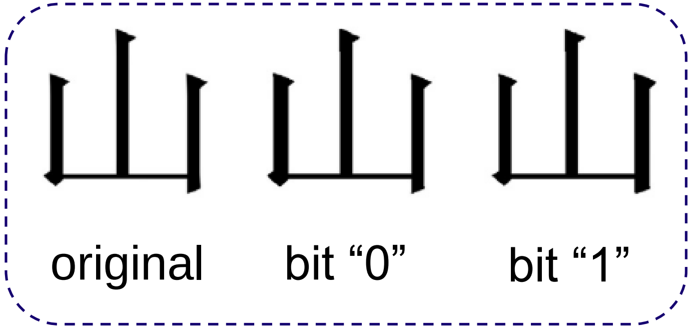
## ❑ Solutions

Non-marked: I *want* to go to Shanghai .

Marked: I *hope* to go to Shanghai .

- ❑ Semantic based
  - ❑ Replace the original word with a word with similar meaning
- ❑ Format based
  - ❑ Adjust the interspace distance between characters, words or paragraphs
  - ❑ Change the color of characters or background and so on
- ❑ **Font based**
  - ❑ Modify the structures of glyph in fonts


0 0 0 1 0 0 0
The methods based on text spacing proposed ...
The methods based on text spacing proposed ...



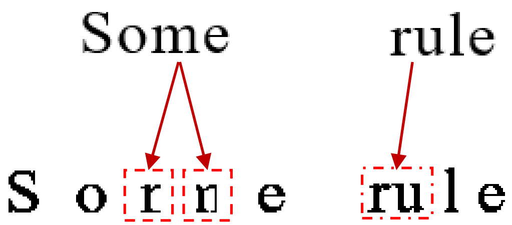
# 1 Motivation

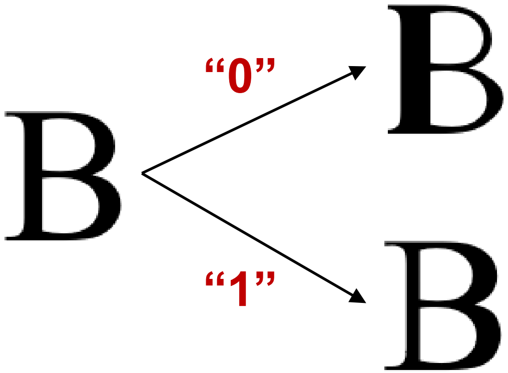
## Challenges (Font based)

- Existing works cannot well adapt to small font sizes
  - High bit error rate (BER) when the font size is small, e.g., < 12 pt
  - Segmentation errors arise due to the small word interspace
- Existing works easily introduce noticeable distortion
  - The original glyph will be modified to carry either bit “0” or “1”
  - The modification intensity (for watermark embedding) is strong

**Screenshots:**

**Segmentation errors:**





# Outline

1 Motivation

2 Proposed Method

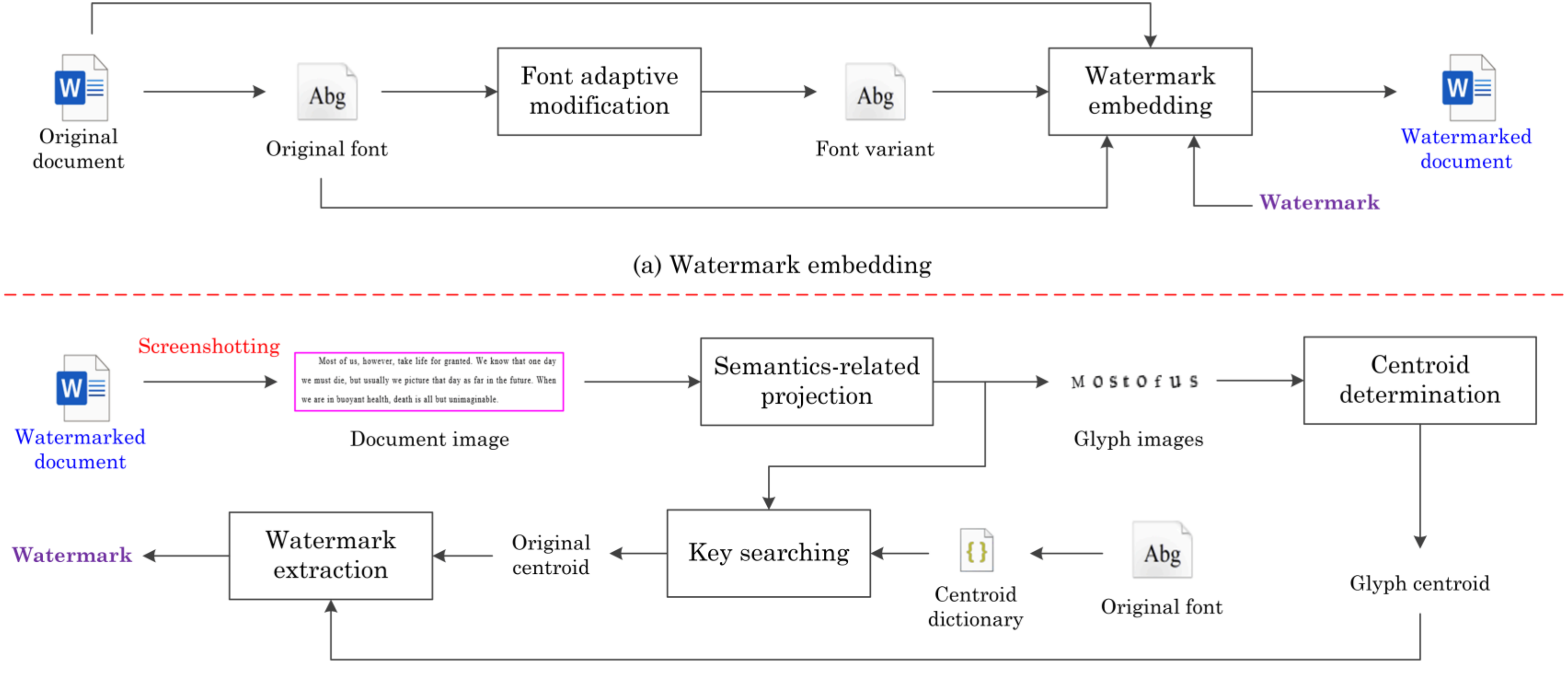
3 Experimental Results and Analysis

4 Conclusion and Discussion

# 2 Proposed Method

## ❑ General Framework

### ❑ *Font adaptive modification & Semantics-related segmentation*



Most of us, however, take life for granted. We know that one day we must die, but usually we picture that day as far in the future. When we are in buoyant health, death is all but unimaginable.

(a) Watermark embedding

(b) Watermark extraction

# 2 Proposed Method

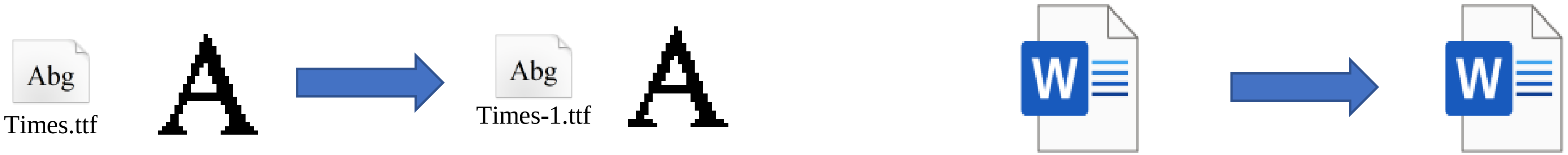
## ❑ Watermark Embedding

### ❑ Font adaptive modification

- ➢ The glyph of the original font is used to carry secret bit “0”
- ➢ The centroid of the glyph of the original font is shifted to carry secret bit “1”

### ❑ Watermark bits embedding

- ➢ Use the original glyph and the marked glyph for watermark embedding



The original font

**The modified font**

Original document

**Watermarked document**

# 2 Proposed Method

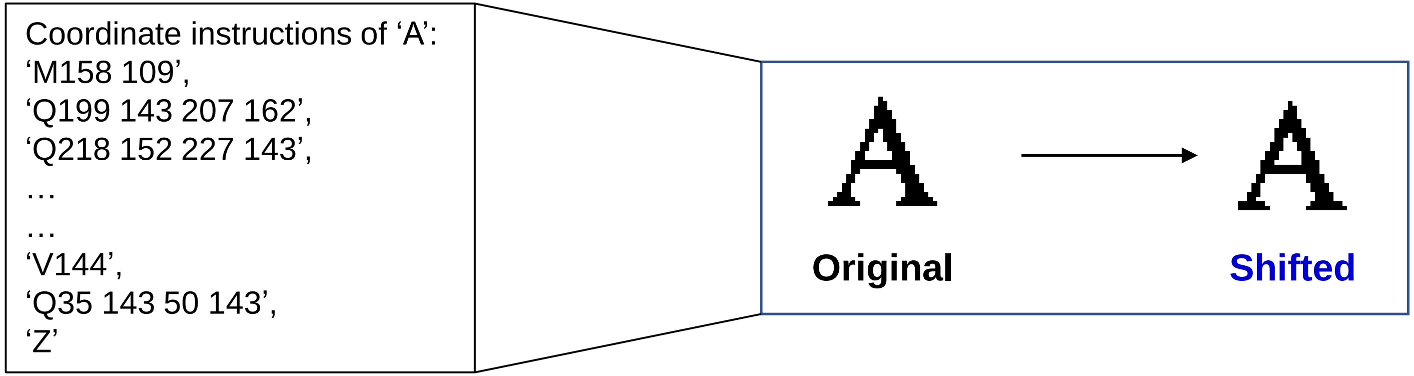
## □ *Glyph Centroid Modification*

### □ Step 1: Calculate the centroid of the glyph of the original font

- ➢ Shift the centroid to ***right*** if the centroid lies on the left side
- ➢ Shift the centroid to ***left*** if the centroid lines on the right side

### □ Step 2: Adjust coordinates to match centroid modification

- ➢ Modify stroke positions and thickness of the glyph

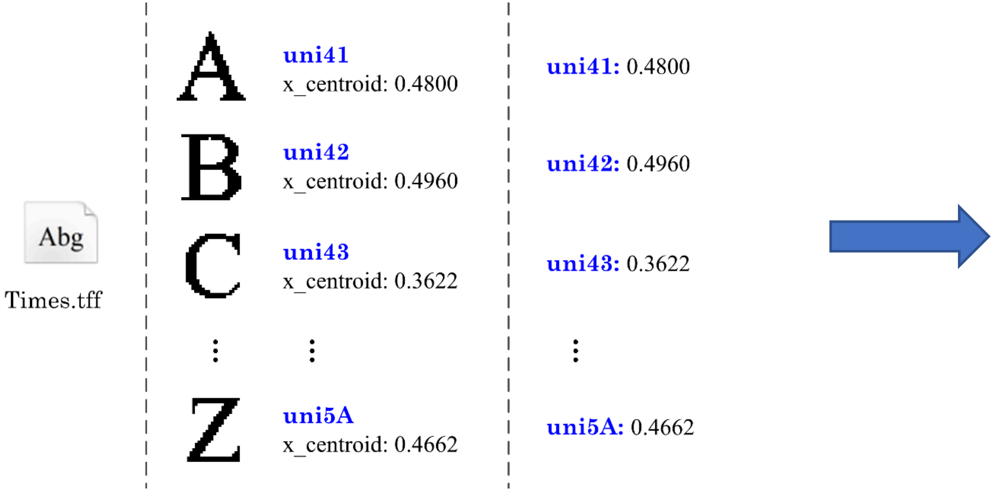


Coordinate instructions of ‘A’:
‘M158 109’,
‘Q199 143 207 162’,
‘Q218 152 227 143’,
…
…
‘V144’,
‘Q35 143 50 143’,
‘Z’

# 2 Proposed Method

## ❑ *Centroid Dictionary*

- ❑ Store the centroids of the original glyphs
- ❑ Centroid dictionary generation
  - ➢ Step 1: Generate the original glyph images and calculate their centroids
  - ➢ Step 2: Save each glyph Unicode and its centroid as a key-value pair



uni41
x_centroid: 0.4800

uni42
x_centroid: 0.4960

uni43
x_centroid: 0.3622

⋮

uni5A
x_centroid: 0.4662

**uni41:** 0.4800
**uni42:** 0.4960
**uni43:** 0.3622
⋮
**uni5A:** 0.4662

*Used for watermark extraction*

# 2 Proposed Method

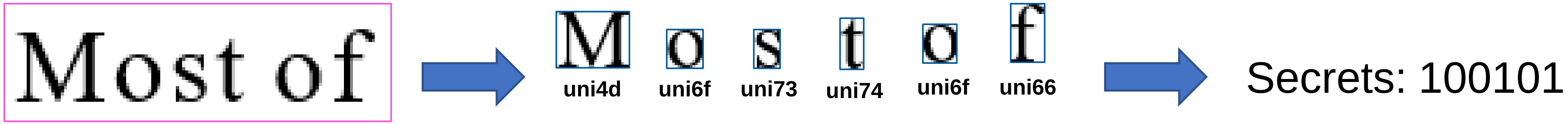
## ❑ Watermark Extraction

### ❑ Semantics-related segmentation

- Recognition: to identify semantics and perform rough segmentation
- Projection: to remove redundant pixels and obtain precise glyph images

### ❑ Watermark bits extraction

- Compare the centroid with dictionary to extract the watermark bit



Secrets: 100101

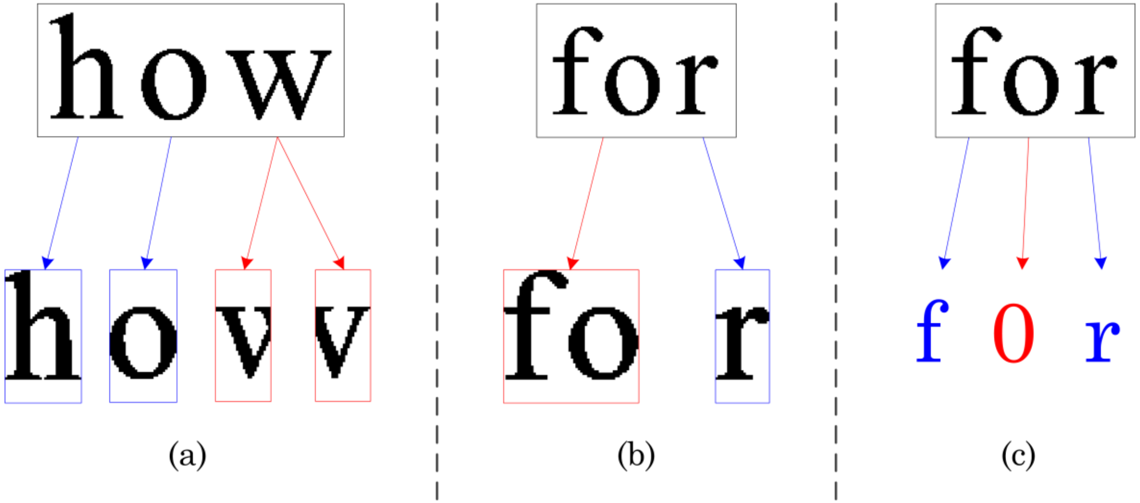
Screenshot image

Glyphs images

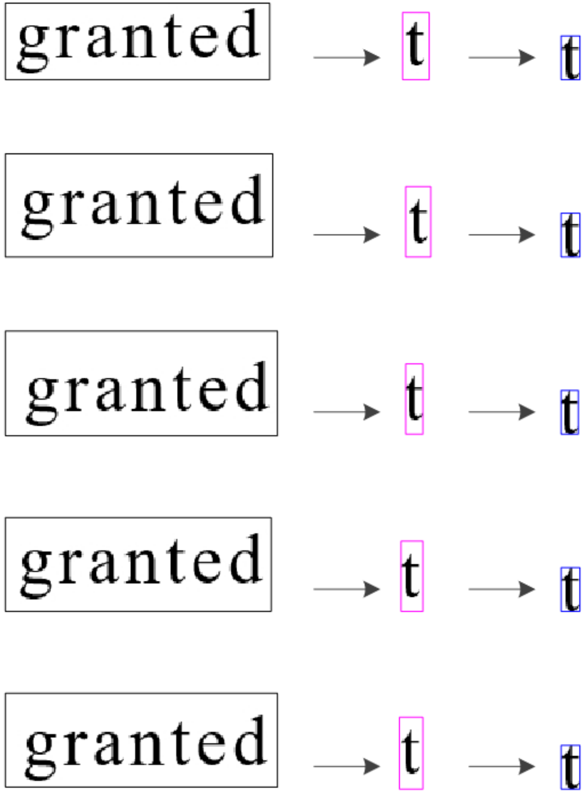
# 2 Proposed Method

## Semantics-related Segmentation

- Reduce segmentation errors
- Obtain stable outputs



Typical segmentation errors



Different screenshots result in stable outputs

# Outline

1 Motivation

2 Proposed Method

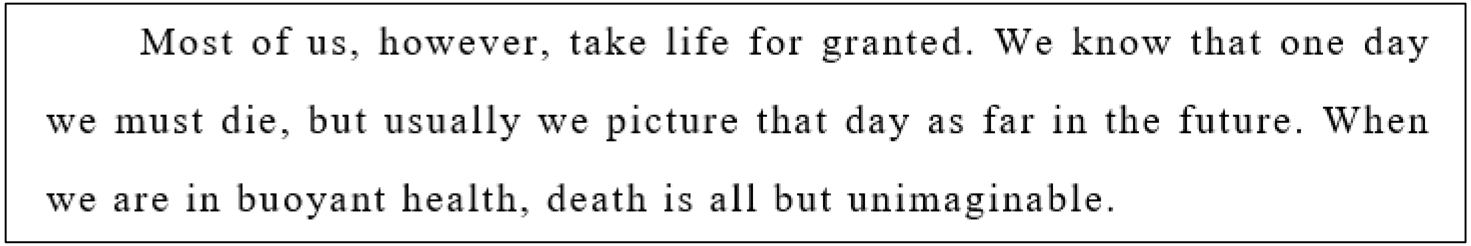
3 Experimental Results and Analysis

4 Conclusion and Discussion

# 3 Experimental Results and Analysis

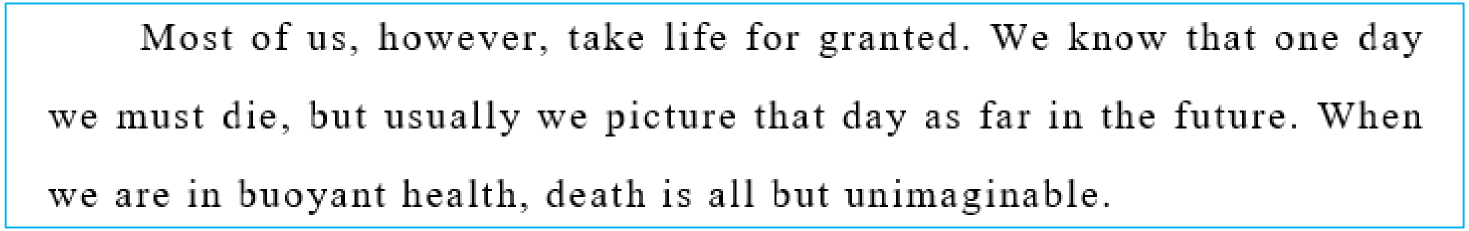
## Qualitative Results

- Satisfactory visual quality & outperform related works



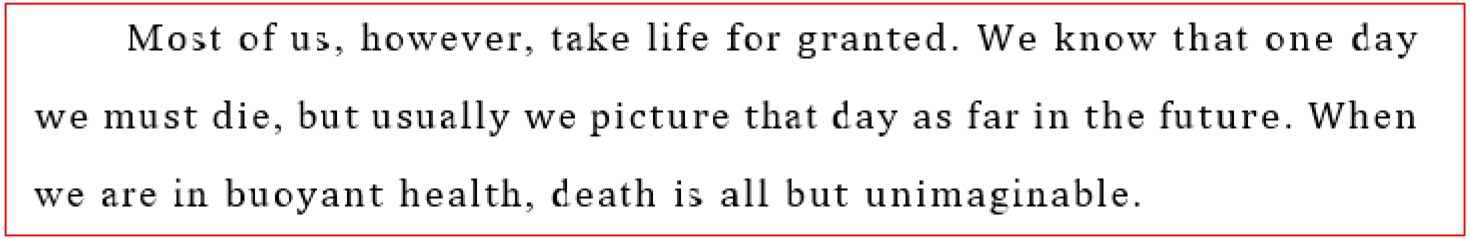
Most of us, however, take life for granted. We know that one day we must die, but usually we picture that day as far in the future. When we are in buoyant health, death is all but unimaginable.

**Original docx**



Most of us, however, take life for granted. We know that one day we must die, but usually we picture that day as far in the future. When we are in buoyant health, death is all but unimaginable.

**Proposed**



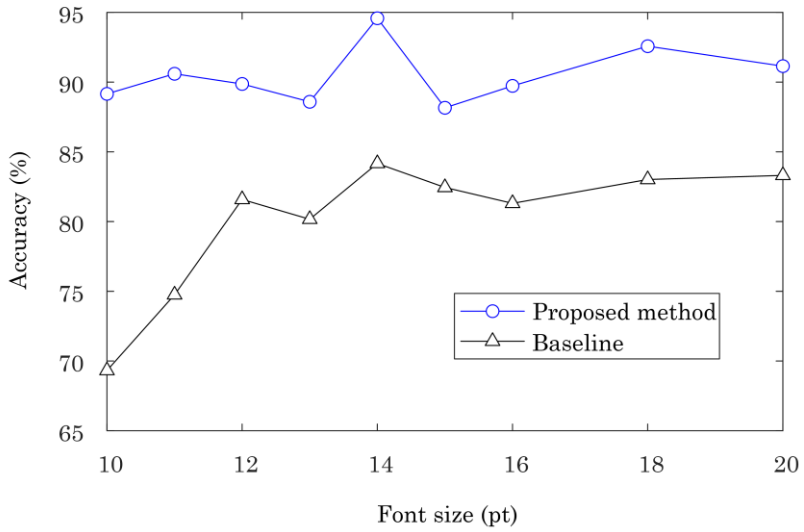
Most of us, however, take life for granted. We know that one day we must die, but usually we picture that day as far in the future. When we are in buoyant health, death is all but unimaginable.

**Baseline**

# 3 Experimental Results and Analysis

## Quantitative Results

- English texts with different font sizes
- Experimental settings
    - Font: Times New Roman
    - Size: 10 ~ 20 pt
    - Content: from a novel
    - Number of chars: ~ 700
- 5%~10% higher than baseline averagely
- No letter was incorrectly segmented



*Accuracy*

*Incorrect glyph segmentations*

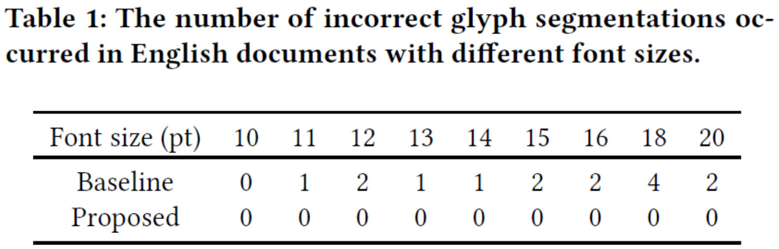


**Table 1: The number of incorrect glyph segmentations occurred in English documents with different font sizes.**

| Font size (pt) | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| Baseline | 0 | 1 | 2 | 1 | 1 | 2 | 2 | 4 | 2 |
| Proposed | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# 3 Experimental Results and Analysis

## Quantitative Results

- Chinese texts with different font sizes
- Experimental settings
  - Font: Simhei
  - Size: 10 ~ 20 pt
  - Content: from news webpages
  - Number of chars: ~ 350
- For languages with fixed width and height in glyphs, the watermarking performance performs better due to *semantics-related segmentation*

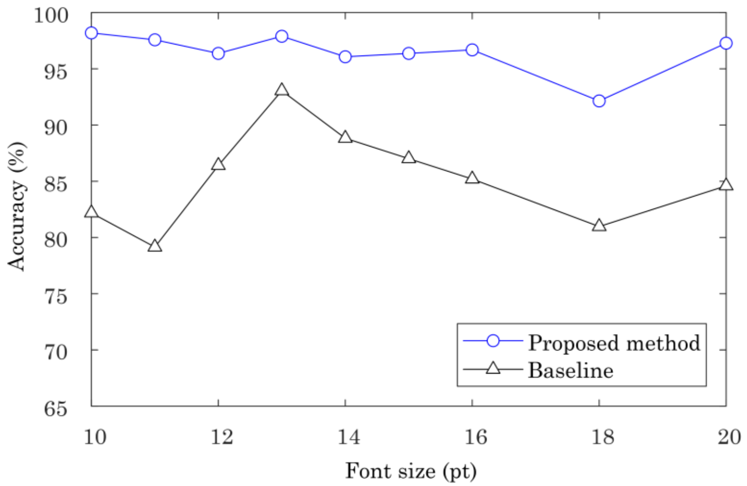
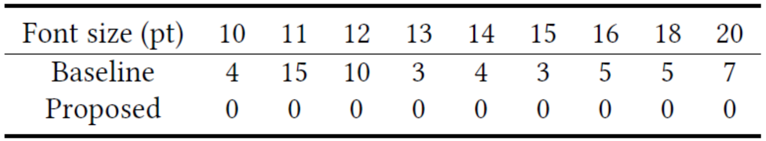


**Table 2: The number of incorrect glyph segmentations occurred in Chinese documents with different font sizes.**

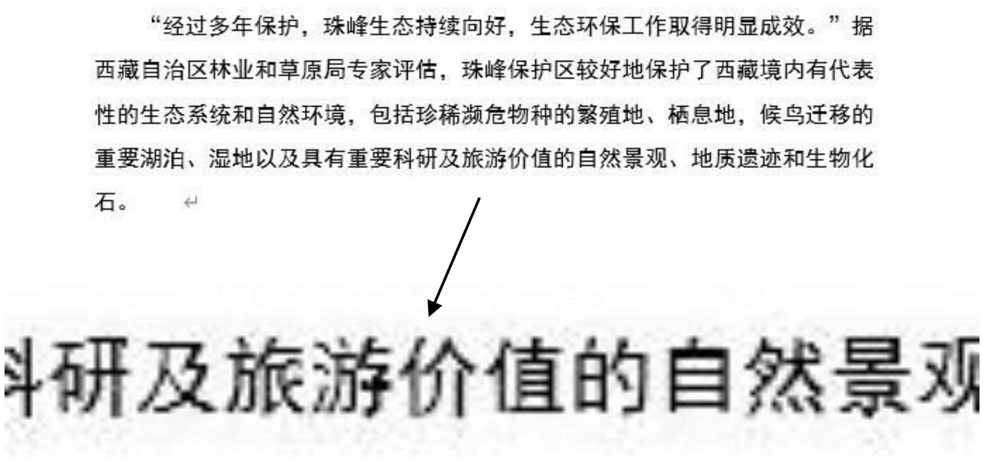


| Font size (pt) | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| Baseline | 4 | 15 | 10 | 3 | 4 | 3 | 5 | 5 | 7 |
| Proposed | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# 3 Experimental Results and Analysis

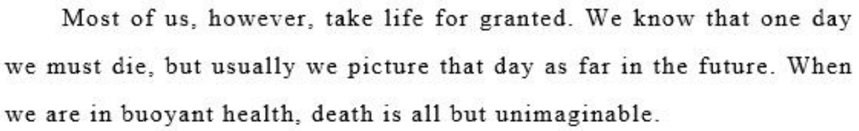
## ❑ Quantitative Results

### ❑ Robust against screenshots after JPEG compression



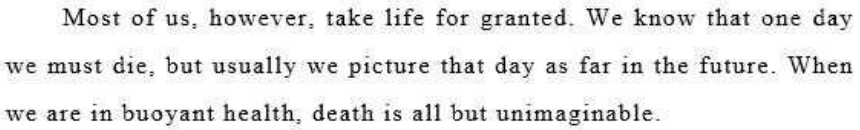
“经过多年保护，珠峰生态持续向好，生态环保工作取得明显成效。”据西藏自治区林业和草原局专家评估，珠峰保护区较好地保护了西藏境内有代表性的生态系统和自然环境，包括珍稀濒危物种的繁殖地、栖息地，候鸟迁移的重要湖泊、湿地以及具有重要科研及旅游价值的自然景观、地质遗迹和生物化石。↵

JPEG compression artifacts



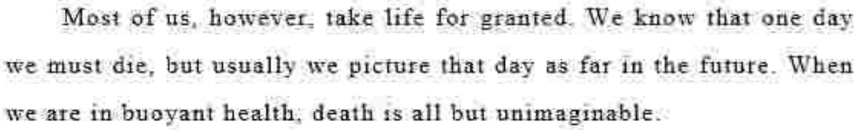
Most of us, however, take life for granted. We know that one day we must die, but usually we picture that day as far in the future. When we are in buoyant health, death is all but unimaginable.

JPEG with 100% compression rate



Most of us, however, take life for granted. We know that one day we must die, but usually we picture that day as far in the future. When we are in buoyant health, death is all but unimaginable.

JPEG with 50% compression rate



Most of us, however, take life for granted. We know that one day we must die, but usually we picture that day as far in the future. When we are in buoyant health, death is all but unimaginable.

JPEG with 10% compression rate

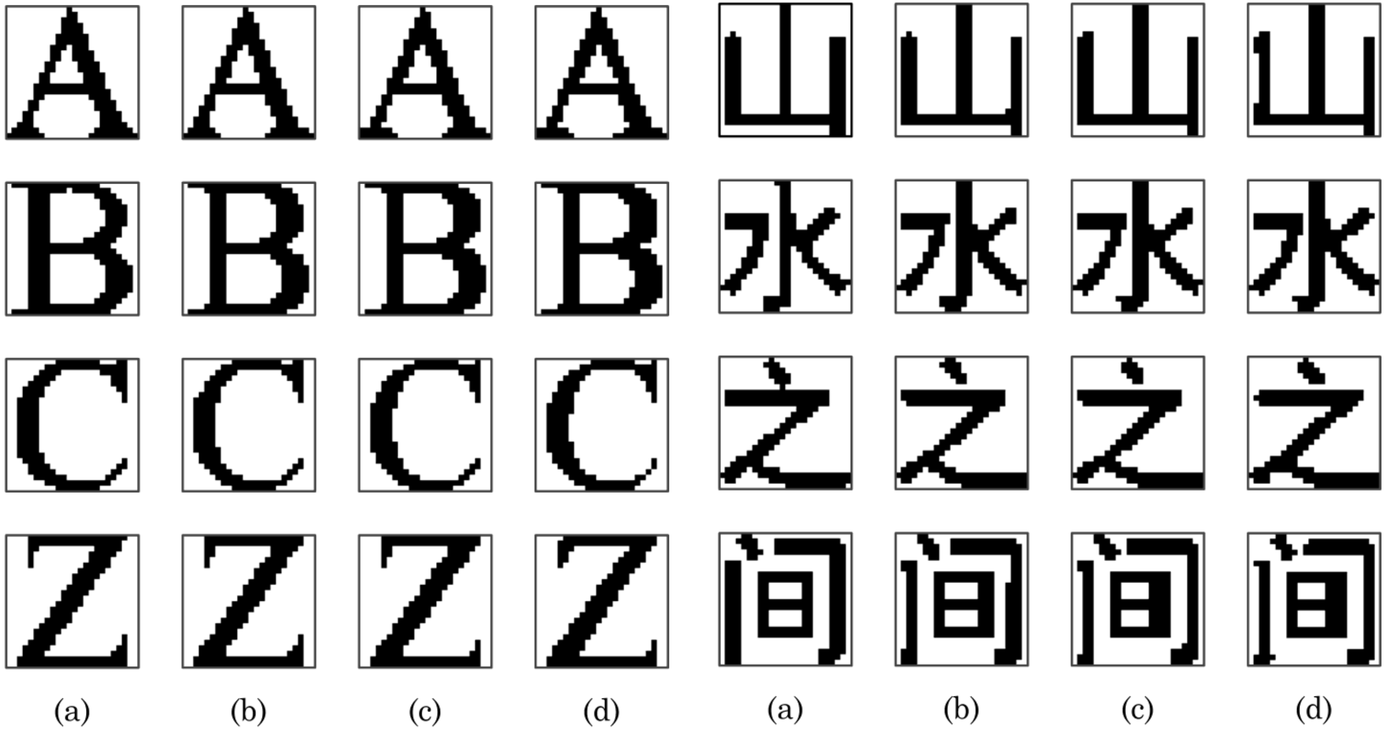
| Format | PNG | JEPG-10 | JEPG-50 | JPEG-100 |
|---|---|---|---|---|
| **Accuracy** | 96.37% | 90.33% | 92.15% | 93.05% |

| Format | PNG | JPEG-10 | JPEG-50 | JPEG-100 |
|---|---|---|---|---|
| **Accuracy** | 93.96% | 87.25% | 89.93% | 91.95% |

# 3 Experimental Results and Analysis

## ❑ Ablation Study

### ❑ Different modification strengths



*α: a system parameter controlling the modification strength*

| Language | $\alpha = 1/24$ | $\alpha = 1/16$ | $\alpha = 1/12$ |
|---|---|---|---|
| Chinese | 87.61% | 96.07% | 96.68% |
| English | 84.31% | 89.87% | 90.16% |

*The larger α, the larger the strength*

*The larger α, the larger the distortion*

*The larger α, the larger the accuracy*

上海大学 SHANGHAI UNIVERSITY

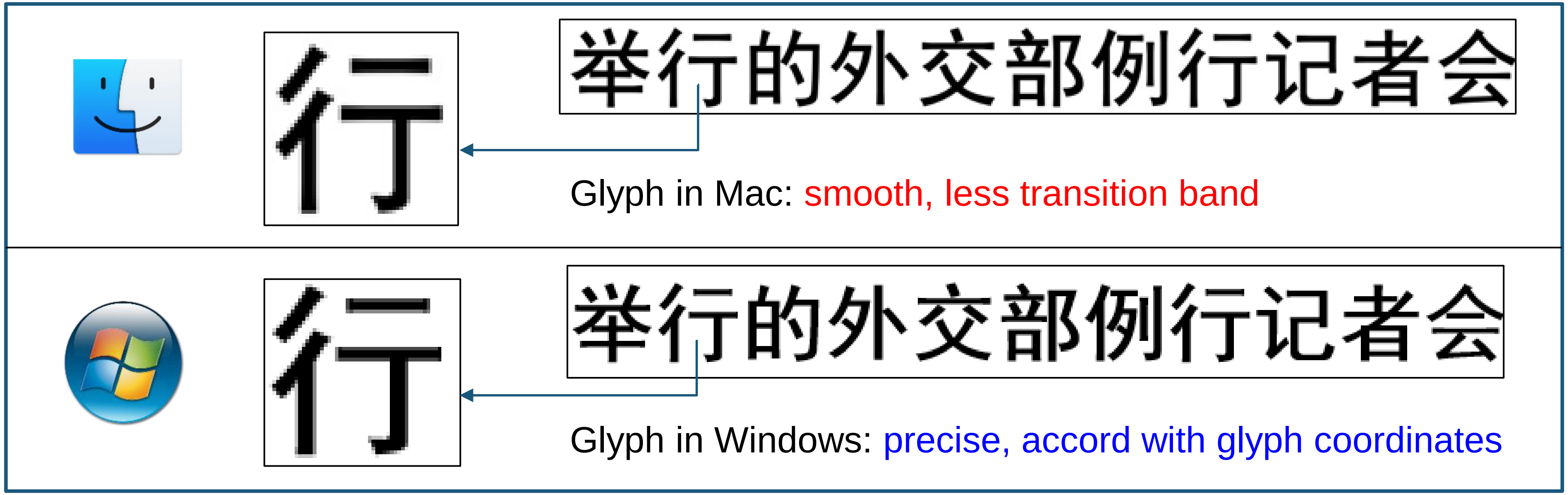
# 3 Experimental Results and Analysis

## ❑ Ablation Study

| Language | Windows | MacOS |
|---|---|---|
| Chinese | 92.86% | 90.18% |
| English | 89.93% | 87.91% |

### ❑ Different operating systems

- ❑ Different font rendering engines: *determine how to display font on screen*
- ❑ The watermark extraction accuracy remains at a high level



Glyph in Mac: smooth, less transition band

Glyph in Windows: precise, accord with glyph coordinates

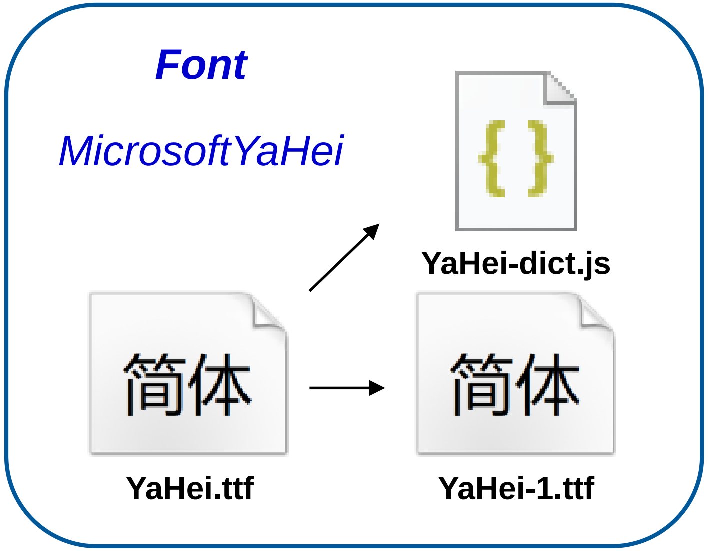
# 3 Experimental Results and Analysis

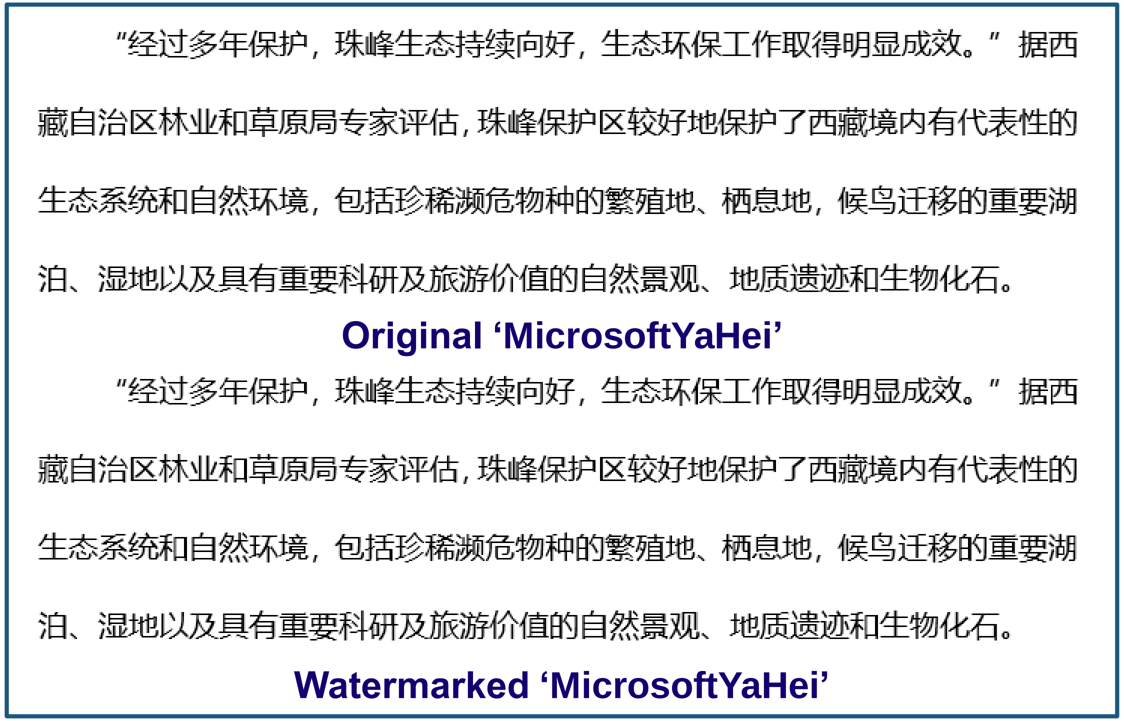
## ❑ Ablation Study

### ❑ Different font styles

| Font | Text 1 | Text 2 | Text 3 |
|---|---|---|---|
| **‘Simhei’** | 94.37% | 95.29% | 91.92% |
| **‘Simsun’** | 92.96% | 94.12% | 89.90% |
| **‘MicrosoftYaHei’** | 91.55% | 91.76% | 90.91% |

The proposed work is not subjected to ***any font styles***





“经过多年保护，珠峰生态持续向好，生态环保工作取得明显成效。” 据西藏自治区林业和草原局专家评估，珠峰保护区较好地保护了西藏境内有代表性的生态系统和自然环境，包括珍稀濒危物种的繁殖地、栖息地，候鸟迁移的重要湖泊、湿地以及具有重要科研及旅游价值的自然景观、地质遗迹和生物化石。

**Original ‘MicrosoftYaHei’**

“经过多年保护，珠峰生态持续向好，生态环保工作取得明显成效。” 据西藏自治区林业和草原局专家评估，珠峰保护区较好地保护了西藏境内有代表性的生态系统和自然环境，包括珍稀濒危物种的繁殖地、栖息地，候鸟迁移的重要湖泊、湿地以及具有重要科研及旅游价值的自然景观、地质遗迹和生物化石。

**Watermarked ‘MicrosoftYaHei’**

# 4 Conclusion and Discussion

## Conclusion

- Apply font adaptive modification and semantics-related segmentation for robustness enhancement of the watermark
- Robust against screenshot (plus JPEG compression)

## Discussion

- Application scenarios: computer screenshot, camera shooting, JPEG compression (plus other potential attacks)
- Still long way to go …

# Many Thanks!