1st Place Solution for ICDAR 2021 Competition on Mathematical Formula Detection

Xianbiao Qi (齐宪标)

Ping An Property and Casualty Insurance Company of China[1]

July 22, 2021

中国平安

金融 · 科技

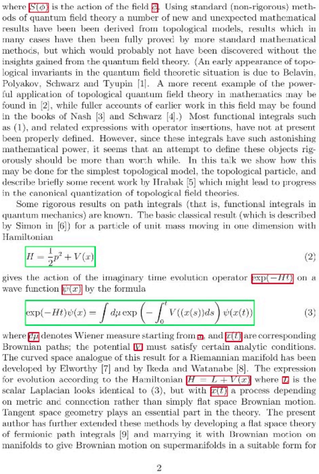
[1] Yuxiang Zhong, Xianbiao Qi, Shanjun Li, Dengyi Gu, Yihao Chen, Peiyang Ning, Rong Xiao

Outline

- Introduction to Mathematical Formula Detection
- Major challenges
- Key points of our solution[2]
- Experiment

[2] Yuxiang Zhong, et al.: 1st Place Solution for ICDAR 2021 Competition on Mathematical Formula Detection. CoRR abs/2107.05534 (2021)

MFD Task



where $S(\phi)$ is the action of the field ϕ. Using standard (non-rigorous) methods of quantum field theory a number of new and unexpected mathematical results have been been derived from topological models, results which in many cases have then been fully proved by more standard mathematical methods, but which would probably not have been discovered without the insights gained from the quantum field theory. (An early appearance of topological invariants in the quantum field theoretic situation is due to Belavin, Polyakov, Schwarz and Tyupin [1]. A more recent example of the powerful application of topological quantum field theory in mathematics may be found in [2], while fuller accounts of earlier work in this field may be found in the books of Nash [3] and Schwarz [4].) Most functional integrals such as (1), and related expressions with operator insertions, have not at present been properly defined. However, since these integrals have such astonishing mathematical power, it seems that an attempt to define these objects rigorously should be more than worth while. In this talk we show how this may be done for the simplest topological model, the topological particle, and describe briefly some recent work by Hrabak [5] which might lead to progress in the canonical quantization of topological field theories.

Some rigorous results on path integrals (that is, functional integrals in quantum mechanics) are known. The basic classical result (which is described by Simon in [6]) for a particle of unit mass moving in one dimension with Hamiltonian

$$H = \frac{1}{2}p^2 + V(x) \tag{2}$$

gives the action of the imaginary time evolution operator $\exp(-Ht)$ on a wave function $\psi(x)$ by the formula

$$\exp(-Ht)\psi(x) = \int d\mu \exp\left(-\int_0^t V((x(s))ds\right)\psi(x(t)) \tag{3}$$

where $d\mu$ denotes Wiener measure starting from x, and $x(t)$ are corresponding Brownian paths; the potential V must satisfy certain analytic conditions. The curved space analogue of this result for a Riemannian manifold has been developed by Elworthy [7] and by Ikeda and Watanabe [8]. The expression for evolution according to the Hamiltonian $H = L + V(x)$ where L is the scalar Laplacian looks identical to (3), but with $x(t)$ a process depending on metric and connection rather than simply flat space Brownian motion. Tangent space geometry plays an essential part in the theory. The present author has further extended these methods by developing a flat space theory of fermionic path integrals [9] and marrying it with Brownian motion on manifolds to give Brownian motion on supermanifolds in a suitable form for

2

Isolated and Embedded Formulas.

```
# ====================================
# class => {0: embedded, 1: isolated}
# ====================================
# x_rel  y_rel  width  height  class
  22.18  53.76  14.58  3.22    1
  22.18  62.21  47.55  3.86    1
  22.94  15.77  4.01   1.46    0
  48.45  15.82  1.24   1.37    0
  69.18  58.15  8.64   1.51    0
  29.79  59.86  4.08   1.51    0
  63.72  67.24  3.39   1.46    0
  22.74  67.29  2.28   1.37    0
  57.98  67.68  1.11   0.68    0
  44.78  69.04  1.66   1.03    0
  55.70  74.07  13.27  1.51    0
  75.54  74.17  1.38   1.07    0
  60.68  75.83  3.39   1.46    0
```

Ground Truth.

Samples

where V is the original rank n vector bundle over X. Second, we must specify a line bundle ℓ on the surface $\pi^* z$ which we will glue together with N to produce $\tilde{N}$. A priori, ℓ is arbitrary, but, as we will see, it is actually subject to strong constraints. To find these constraints, we first use the Fourier–Mukai transformation to construct a vector bundle V_z from the spectral data $(\pi^* z, \ell)$. That is,

$$(\pi^* z, \ell) \longrightarrow V_z \tag{5.5}$$

where V_z is a rank 1 vector bundle over the curve $\tilde{z} = \sigma \cdot \pi^* z$.[2] Now, the fact that, by construction, the base component of the bulk five–brane class associated with $\tilde{V}$ must be equal to $W_B - \pi^* z$, implies that $\ell = \pi^* L$ for some line bundle L on the curve z in the base. A simple Fourier-Mukai calculation shows that V_z is just

$$V_z = i_*(L \otimes K_B) \tag{5.6}$$

where i is the embedding $i : \tilde{z} \to X$ of the curve $\tilde{z}$ in X.

Given V and V_z, we now attempt to "weave" them together by relating them on the curve $\tilde{z}$ in X where their data overlap. This is done by specifying a surjection

$$\xi : V|_{\tilde{z}} \longrightarrow V_z \tag{5.7}$$

That such a surjection exists and is unique will become clear shortly. Given this relation, one can define a "bundle" $\widetilde{V}$ on X via the exact sequence[3]

$$0 \to \widetilde{V} \to V \to V_z \to 0 \tag{5.8}$$

It can be shown that because

$$\text{codimension } \tilde{z} = 2 > 1, \tag{5.9}$$

then $\widetilde{V}$ is a singular object, called a torsion free sheaf[4], but is not a smooth vector bundle. Hence, to complete our construction, we will have to show that $\widetilde{V}$ can be "smoothed out" to

[2] As a technical aside, we note that the vector bundle V_z over the curve $\tilde{z}$ can be formally extended over the Calabi–Yau threefold X as an object that vanishes everywhere outside the curve $\tilde{z}$ and is identical to V_z on $\tilde{z}$. This extended object will also be denoted by V_z. We will let context dictate which of these notions is to be used. This remark applies to all of the line bundles discussed throughout this paper.

[3] The bundle $\widetilde{V}$ defined in this way bears a special name. It is a called a Hecke transform of V and the pair (ξ, V_z) is called the center of the Hecke transform.

[4] The notion of stability here is similar to that used for vector bundles. The differential geometric counterpart of a stable sheaf $\widetilde{V}$ is a Hermitian-Yang-Mills connection A on the vector bundle $\widetilde{V}$ which is smooth outside of the curve $\tilde{z} \subset X$ but has a delta function behavior along the curve $\tilde{z}$. In other words, a stable torsion free sheaf $\widetilde{V}$ is the algebraic-geometry incarnation of a "small instanton" concentrated on the curve $\tilde{z}$. This justifies the terminology "small instanton phase transition"that we use below to describe the two step process of first creating $\widetilde{V}$ out of V and then deforming $\widetilde{V}$ to a smooth vector bundle V' on X.

15

5 *COINCIDENCE OF THE SPECTRAL DATA* 20

the Hitchin spectral curve $\mathbf{C}$, then the kernel of the map $(z-\zeta)$ from $H^1(K_\sigma)$ to $H^1(K_\sigma)$ is nontrivial. But the cohomology exact sequence implies an isomorphism

$$\mathrm{Ker}(z-\zeta) \cong H^0_{\bar{\mathcal{D}}}(\mathbb{S}^1, E|_{z=\zeta}). \tag{32}$$

Therefore $H^0_{\bar{\mathcal{D}}}(\mathbb{S}^1, E|_{z=\zeta})$ is nontrivial as well. This means that the holonomy of B along the circle $z=\zeta$ has $\exp(2\pi\sigma)$ as one of its eigenvalues. Thus the point $(\zeta, \exp(2\pi\sigma)) \in \mathbb{C} \times \mathbb{C}^*$ belongs to the monopole spectral curve $\mathbf{S}$. Moreover, the fibers of the spectral line bundles on $\mathbf{C}$ and $\mathbf{S}$ are given by $\mathrm{Ker}(z-\zeta)$ and $H^0_{\bar{\mathcal{D}}}(\mathbb{S}^1, E|_{z=\zeta})$, respectively. Thus we also get an isomorphism of the line bundles.

5.4 Exactness Of The Cohomology Sequence

Consider again the complex of sheaves

$$0 \to E \xrightarrow{(z-\zeta)} E \xrightarrow{rest.} E|_{z=\zeta} \to 0, \tag{33}$$

where $rest.$ is the restriction to $z=\zeta$. Since $\bar{\mathcal{D}}_p,\ p=0,1,$ commutes with $(z-\zeta)$ and $rest.$, this complex is included in a double complex $D^{p,q}$:

$$D^{p,q}: \quad \begin{array}{ccccc} E\otimes\Lambda^{0,2} & \xrightarrow{(z-\zeta)} & E\otimes\Lambda^{0,2} & \xrightarrow{rest.} & 0 \\ \bar{\mathcal{D}}_1 \uparrow & & \bar{\mathcal{D}}_1 \uparrow & & \bar{\mathcal{D}}_1 \uparrow \\ E\otimes\Lambda^{0,1} & \xrightarrow{(z-\zeta)} & E\otimes\Lambda^{0,1} & \xrightarrow{rest.} & E\otimes\Lambda^{0,1}|_{z=\zeta} \\ \bar{\mathcal{D}}_0 \uparrow & & \bar{\mathcal{D}}_0 \uparrow & & \bar{\mathcal{D}}_0 \uparrow \\ E\otimes\Lambda^{0,0} & \xrightarrow{(z-\zeta)} & E\otimes\Lambda^{0,0} & \xrightarrow{rest.} & E\otimes\Lambda^{0,0}|_{z=\zeta}. \end{array}$$

Computing the cohomology of the rows, we obtain the first term of the "vertical" spectral sequence :

$$E_1^{p,q}: \quad \begin{array}{ccc} 0 & \{\eta^{0,2} \sim \eta^{0,2} + (z-\zeta)\omega^{0,2}\} & 0 \\ \uparrow & \bar{\mathcal{D}}_1 \uparrow & \uparrow \\ 0 & \{\eta^{0,1} \sim \eta^{0,1} + (z-\zeta)\omega^{0,1} |\ rest.(\eta^{0,1}) = 0\} & 0 \\ \uparrow & \bar{\mathcal{D}}_0 \uparrow & \uparrow \\ 0 & \{\eta^{0,0} \sim \eta^{0,0} + (z-\zeta)\omega^{0,0} |\ \eta^{0,0}|_{z=\zeta} = 0\} & 0 \end{array}$$

On the second level the "vertical" spectral sequence degenerates to zero: $E_\infty^{p,q} = 0$.

Major Challenges

- Large scale span
- Large variation of the ratio between height and width
- Rich character set and mathematical expressions

Key points of our solution

- Generalized Focal Loss (GFL)
- Adaptive Training Sampling Strategy (ATSS)
- Feature Pyramid Network (FPN)

GFL

Comparison between anchor-based and anchor-free methods

- Pre-set anchors vs no anchors
- Pre-set scale and ratio vs any points on object for positive sampling
- Regression of the box residual vs direct regression

GFL

Drawbacks of anchor-based method on MFD task

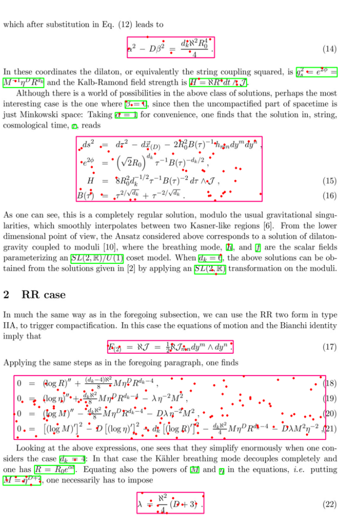
- It is hard to preset effective anchors to cover large span of scales and aspect ratios.
- Even with ATSS, small objects cannot be assigned positive anchors because anchor candidate will be filtered out by IoU.

GFL

- Quality Focal Loss (QFL)
- Distribution Focal Loss (DFL)

ATSS

Random sampling vs ATSS sampling.



which after substitution in Eq. (12) leads to

$$\alpha^2 - D\beta^2 = \frac{d_k \aleph^2 R_0^4}{4}. \tag{14}$$

In these coordinates the dilaton, or equivalently the string coupling squared, is $g_s^2 = e^{2\phi} = M^{-1}\eta^D R^{d_k}$ and the Kalb-Ramond field strength is $H = \aleph R^4 dt \wedge \mathcal{J}$.

Although there is a world of possibilities in the above class of solutions, perhaps the most interesting case is the one where $\beta = 0$, since then the uncompactified part of spacetime is just Minkowski space: Taking $\alpha = 1$ for convenience, one finds that the solution in, string, cosmological time, τ, reads

$$\begin{aligned} ds^2 &= d\tau^2 - d\vec{x}_{(D)}^2 - 2R_0^2 B(\tau)^{-1} h_{mn} dy^m dy^n \,, \\ e^{2\phi} &= \left(\sqrt{2}R_0\right)^{d_k} \tau^{-1} B(\tau)^{-d_k/2} \,, \\ H &= 8R_0^2 d_k^{-1/2} \tau^{-1} B(\tau)^{-2}\, d\tau \wedge \mathcal{J} \,, && (15) \\ B(\tau) &= \tau^{2/\sqrt{d_k}} + \tau^{-2/\sqrt{d_k}} \,. && (16) \end{aligned}$$

As one can see, this is a completely regular solution, modulo the usual gravitational singularities, which smoothly interpolates between two Kasner-like regions [6]. From the lower dimensional point of view, the Ansatz considered above corresponds to a solution of dilaton-gravity coupled to moduli [10], where the breathing mode, R, and $\mathcal{J}$ are the scalar fields parameterizing an $SL(2,\mathbb{R})/U(1)$ coset model. When $d_k = 6$, the above solutions can be obtained from the solutions given in [2] by applying an $SL(2,\mathbb{R})$ transformation on the moduli.

2 RR case

In much the same way as in the foregoing subsection, we can use the RR two form in type IIA, to trigger compactification. In this case the equations of motion and the Bianchi identity imply that

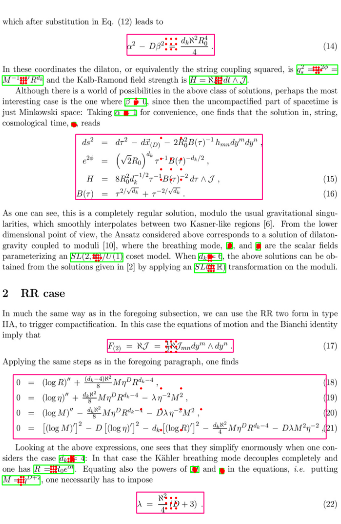
$$F_{(2)} = \aleph \mathcal{J} = \tfrac{1}{2}\aleph \mathcal{J}_{mn} dy^m \wedge dy^n \,. \tag{17}$$

Applying the same steps as in the foregoing paragraph, one finds

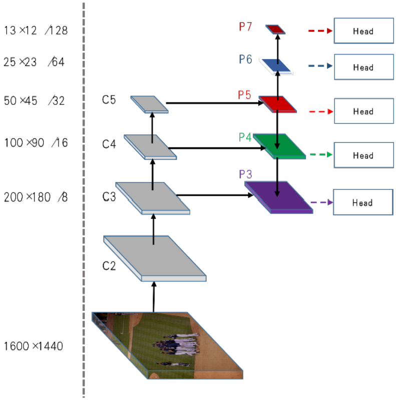
$$\begin{aligned} 0 &= (\log R)'' + \tfrac{(d_k-4)\aleph^2}{8} M\eta^D R^{d_k-4} \,, && (18) \\ 0 &= (\log \eta)'' + \tfrac{d_k\aleph^2}{8} M\eta^D R^{d_k-4} - \lambda\eta^{-2}M^2 \,, && (19) \\ 0 &= (\log M)'' - \tfrac{d_k\aleph^2}{8} M\eta^D R^{d_k-4} - D\lambda\eta^{-2}M^2 \,, && (20) \\ 0 &= [(\log M)']^2 - D[(\log \eta)']^2 - d_k[(\log R)']^2 - \tfrac{d_k\aleph^2}{4} M\eta^D R^{d_k-4} - D\lambda M^2\eta^{-2} \,. && (21) \end{aligned}$$

Looking at the above expressions, one sees that they simplify enormously when one considers the case $d_k = 4$: In that case the Kähler breathing mode decouples completely and one has $R = R_0 e^{\alpha t}$. Equating also the powers of M and η in the equations, *i.e.* putting $M = \eta^{D+2}$, one necessarily has to impose

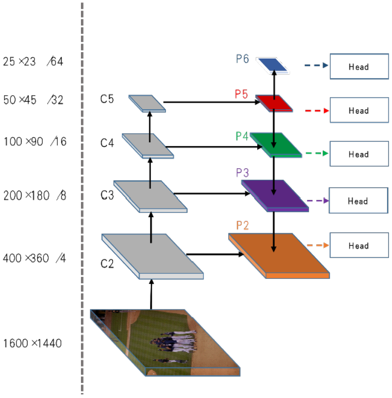
$$\lambda = \frac{\aleph^2}{4}(D+3) \,. \tag{22}$$



FPN



FPN 3-7



FPN 2-6

Tricks

- Double Training Epoch
- Large Crop Size
- Random Flip
- ResNeSt
- Synchronized Batch Normalization
- Deformable Convolution Network
- Larger Batch Size
- Ranger
- RegMax
- FPN Selection
- Weight Box Fusion

Ablation Study

DTE	LCS	Flip	NeSt	SyBN	DCN	LBS	Ranger	Reg 24	FPN (2-6)	WBF	F1-score Embedded	F1-score Isolated	F1-score Total
											89.88 p:91.34 r:88.47	85.92 p:89.77 r:82.39	**89.17** p:91.02 r:87.39
√											91.36 p:92.80 r:89.96	86.29 p:90.93 r:82.10	**90.45** p:92.26 r:88.71
√	√										92.95 p:94.43 r:91.52	87.84 p:92.04 r:84.01	**92.03** p:93.58 r:90.53
√	√	√									93.58 p:94.79 r:92.41	88.38 p:92.29 r:84.78	**92.66** p:94.35 r:91.03
√	√	√	√								93.49 p:93.82 r:93.16	91.33 p:93.56 r:89.20	**93.12** p:93.71 r:92.54
√	√	√	√	√							93.81 p:94.00 r:93.63	91.60 p:93.82 r:89.48	**93.42** p:93.97 r:92.89
√	√	√	√	√	√						94.33 p:95.62 r:93.07	95.25 p:95.62 r:94.88	**94.49** p:95.62 r:93.39
√	√	√	√	√	√	√					94.58 p:95.16 r:94.00	95.60 p:95.97 r:95.23	**94.76** p:95.31 r:94.22
√	√	√	√	√	√	√	√				95.22 p:95.70 r:94.74	95.81 p:96.12 r:95.51	**95.33** p:95.79 r:94.87
√	√	√	√	√	√	√	√	√			95.01 p:95.86 r:94.16	97.28 p:97.18 r:97.38	**95.41** p:96.10 r:94.73
√	√	√	√	√	√	√	√	√	√		95.67 p:96.34 r:95.00	97.67 p:97.46 r:97.88	**96.03** p:96.54 r:95.53
√	√	√	√	√	√	√	√	√	√	√	96.01 p:95.79 r:96.23	98.14 p:97.32 r:98.98	**96.33** p:96.81 r:95.85

Table 1: Experimental results of baseline, DTE (double training epoch), LCS (larger crop size), Flip (training and testing time flip), SyBN (SyncBN), DCN (deformable convolution network), LBS (larger batch size), Ranger optimizer, Reg 24 (a hyperparameter regmax), FPN (2-6) (FPN selection), WBF (weighted box fusion).

Unsuccessful Attempts

- CutMix data augmentation
- Soft NMS vs original NMS
- IoU loss or GIoU loss vs smooth L1 loss
- BFP or PAFPN instead of original FPN
- Swish or Mish vs ReLU

Final Results of 15 Teams

Group ID	Type	F1 score (Ts10 + Ts11)	F1 score Task dependent (Ts11)	F1 score Task independent (Ts10)
PAPCIC	E	94.79 (94.89 , 94.69)	95.11 (95.11 , 95.11)	94.64 (94.79 , 94.50)
	I	98.76 (98.25 , 99.28)	98.70 (98.34 , 99.07)	98.79 (98.21 , 99.37)
	S	**95.47** (95.47 , 95.47)	**95.68** (95.62 , 95.73)	**95.37** (95.40 , 95.35)
Lenovo	E	94.29 (95.36 , 93.25)	93.98 (95.41 , 92.60)	94.44 (95.34 , 93.56)
	I	98.19 (98.26 , 98.12)	97.85 (98.04 , 97.67)	98.33 (98.35 , 98.31)
	S	**94.96** (95.86 , 94.08)	**94.60** (95.84 , 93.40)	**95.13** (95.87 , 94.39)
DLVCLab	E	93.79 (94.54 , 93.05)	93.88 (94.70 , 93.07)	93.75 (94.46 , 93.04)
	I	98.54 (98.19 , 98.89)	98.61 (98.33 , 98.88)	98.51 (98.13 , 98.90)
	S	**94.60** (95.17 , 94.04)	**94.64** (95.29 , 93.99)	**94.59** (95.12 , 94.07)
TYAI	E	93.39 (94.43 , 92.38)	93.94 (95.05 , 92.86)	93.13 (94.13 , 92.15)
	I	98.55 (98.19 , 98.92)	98.42 (98.15 , 98.70)	98.61 (98.21 , 99.02)
	S	**94.28** (95.08 , 93.49)	**94.66** (95.55 , 93.79)	**94.1** (94.86 , 93.35)
SPDBLab	E	92.80 (93.25 , 92.36)	92.14 (92.83 , 91.46)	93.12 (93.45 , 92.79)
	I	98.06 (98.06 , 98.06)	97.76 (97.85 , 97.67)	98.19 (98.15 , 98.23)
	S	**93.70** (94.08 , 93.33)	**93.03** (93.63 , 92.44)	**94.01** (94.28 , 93.74)
YoudaoAI	E	92.73 (93.57 , 91.91)	92.71 (93.95 , 91.51)	92.74 (93.39 , 92.10)
	I	98.34 (97.66 , 99.03)	98.38 (97.97 , 98.79)	98.32 (97.52 , 99.13)
	S	**93.70** (94.28 , 93.12)	**93.63** (94.61 , 92.67)	**93.74** (94.14 , 93.34)
PKUF-MFD	E	91.94 (92.18 , 91.70)	92.32 (92.93 , 91.72)	91.76 (91.82 , 91.69)
	I	96.56 (96.88 , 96.24)	96.87 (97.28 , 96.46)	96.43 (96.72 , 96.15)
	S	**92.72** (92.98 , 92.47)	**93.04** (93.62 , 92.47)	**92.57** (92.68 , 92.47)
HW-L	E	90.53 (91.55 , 89.53)	90.57 (91.82 , 89.35)	90.51 (91.42 , 89.61)
	I	98.94 (98.81 , 99.06)	98.61 (98.33 , 98.88)	99.08 (99.02 , 99.13)
	S	**91.97** (92.81 , 91.15)	**91.86** (92.88 , 90.86)	**92.02** (92.77 , 91.28)
Komachi	E	90.39 (90.92 , 89.86)	89.69 (90.43 , 88.97)	90.72 (91.16 , 90.28)
	I	98.57 (98.27 , 98.87)	98.6 (98.69 , 98.51)	98.55 (98.09 , 99.02)
	S	**91.79** (92.19 , 91.39)	**91.11** (91.75 , 90.48)	**92.10** (92.39 , 91.81)
AIG	E	89.71 (90.18 , 89.25)	89.19 (89.96 , 88.44)	89.95 (90.28 , 89.63)
	I	95.95 (99.00 , 93.09)	96.07 (99.01 , 93.30)	95.90 (99.00 , 93.00)
	S	**90.75** (91.61 , 89.90)	**90.26** (91.34 , 89.21)	**90.97** (91.74 , 90.22)
PKUSG	E	89.10 (90.26 , 87.97)	88.59 (90.23 , 87.00)	89.34 (90.27 , 88.42)
	I	97.96 (97.85 , 98.06)	97.94 (98.31 , 97.58)	97.96 (97.66 , 98.27)
	S	**90.62** (91.58 , 89.68)	**90.09** (91.54 , 88.68)	**90.87** (91.60 , 90.15)
TAL	E	87.87 (88.56 , 87.20)	88.51 (89.12 , 87.92)	87.57 (88.29 , 86.85)
	I	96.85 (96.31 , 97.40)	97.09 (96.33 , 97.86)	96.75 (96.30 , 97.21)
	S	**89.42** (89.91 , 88.93)	**89.89** (90.29 , 89.49)	**89.19** (89.73 , 88.67)
UIT	E	86.04 (85.60 , 86.49)	85.64 (85.62 , 85.65)	86.23 (85.58 , 86.89)
	I	97.05 (95.12 , 99.06)	98.11 (97.08 , 99.16)	96.60 (94.31 , 99.02)
	S	**87.94** (87.26 , 88.63)	**87.63** (87.47 , 87.80)	**88.08** (87.16 , 89.02)
AV-DFKI	E	85.35 (87.37 , 83.41)	84.75 (86.96 , 82.66)	85.63 (87.57 , 83.77)
	I	97.48 (97.12 , 97.84)	97.40 (97.30 , 97.49)	97.52 (97.04 , 97.99)
	S	**87.45** (89.10 , 85.87)	**86.80** (88.67 , 85.01)	**87.76** (89.30 , 86.27)
VH	E	84.25 (83.49 , 85.02)	84.39 (83.76 , 85.02)	84.18 (83.37 , 85.01)
	I	98.59 (98.51 , 98.67)	98.51 (98.42 , 98.60)	98.62 (98.55 , 98.70)
	S	**86.67** (86.01 , 87.34)	**86.61** (86.06 , 87.18)	**86.70** (85.99 , 87.41)


中国平安
PING AN
金融·科技